Information propagation in far-from-equilibrium molecular templating networks is optimised by pseudo-equilibrium systems with negligible dissipation

Benjamin Qureshi and Thomas E. Ouldridge*
Department of Bioengineering and Centre for Synthetic Biology, Imperial College London, London SW7 2AZ, United Kingdom

Jenny M. Poulton
Department of Physics and Astronomy, University of Sheffield, Sheffield S3 7RH, United Kingdom

(Dated: April 4, 2024)

Far-from equilibrium molecular templating networks, like those that maintain the populations of RNA and protein molecules in the cell, are key biological motifs. These networks share the general property that assembled products are produced and degraded via complex pathways controlled by catalysts, including molecular templates. Although it has been suggested that the information propagated from templates to products sets a lower bound on the thermodynamic cost of these networks, this bound has not been explored rigorously to date. We show that, for an arbitrarily catalytic reaction network in steady state, the specificity with which a single product can dominate the ensemble is upper bounded, and the entropy of the product ensemble lower bounded, by a function of ΔG, the difference between the maximal and minimal free-energy changes along pathways to assembly. These simple bounds are particularly restrictive for systems with a smaller number of possible products M. Remarkably, however, although ΔG constrains the information propagated to the product distribution, the systems that saturate the bound operate in a pseudo-equilibrium fashion, and there is no minimal entropy production rate for maintaining this non-equilibrium distribution. Moreover, for large systems, a vanishingly small subset of the possible products can dominate the product ensemble even for small values of $\Delta G/\ln M$.

I. INTRODUCTION

Biochemical systems use catalysts to selectively produce multiple distinct products from a shared set of ingredients. For example, consider protein production [1]: many polypeptide sequences are created from the same set of amino acids, using a set of template mRNAs and other molecules act as selective catalysts [2]. The result is a distribution of polypeptide sequences in the cell, sharply peaked around the target sequences specified by the mRNA. Since these sequences fold into functional proteins, a sharply-peaked, or accurate, distribution is essential to biological function.

Often, biological systems continuously turn over the products, recycling the basic ingredients for subsequent use. For example, polynucleotide phosphorylase enzymes actively degrade the pool of selectively transcribed RNA [3]. Degradation allows the cell to set up a finite steady-state population that is biased towards specific products, and continuous turnover and recycling of components is necessary for systems with finite resources that need to change their state in response to dynamic environments. The underlying processes may also be complex and involve many possible routes of assembling and disassembling the products. For example, kinetic proofreading motifs [4–6] introduce fuel-consuming loops into the assembly process, thereby providing multiple pathways to each product.

Previous studies [6–30] have generally focused on models of a single production process only, in particular considering how templated copolymerisation in isolation can produce an accurate sequence. These studies have explored thermodynamic constraints on the accuracy in different classes of templated copolymerisation models, including those with single pathways leading to infinite-length polymers [23–25], more complex infinite-length polymerisation processes [26–28] and finite length copolymerisation [29, 30].

To the best of our knowledge, only Refs. [2, 31, 32] have considered the implications of two or more pathways for production and destruction of an ensemble of products, in each case without considering a detailed model of dynamical pathways. These works hypothesise a minimal thermodynamic cost or entropy production per product made, which is related to the information accurately transferred from template to products [2, 31, 32]. However, entropy production in stochastic thermodynamics is related to the relative rate of forwards and backwards transitions [33], rather than the relative rate of two different transitions; it therefore places no minimal cost on the specificity of a catalyst [29]. Exactly how these two ideas are reconciled is unclear.

The limiting thermodynamics of this crucial class of processes therefore remains under-explored. In this paper, we consider the thermodynamics of a general system in which a distribution of products is created and maintained by a set of catalysts that inter-convert molecules between pools of monomer inputs and pools of assembled products. These inter-conversion processes can be arbitrarily complex, but we assume that all products have

* t.ouldridge@imperial.ac.uk

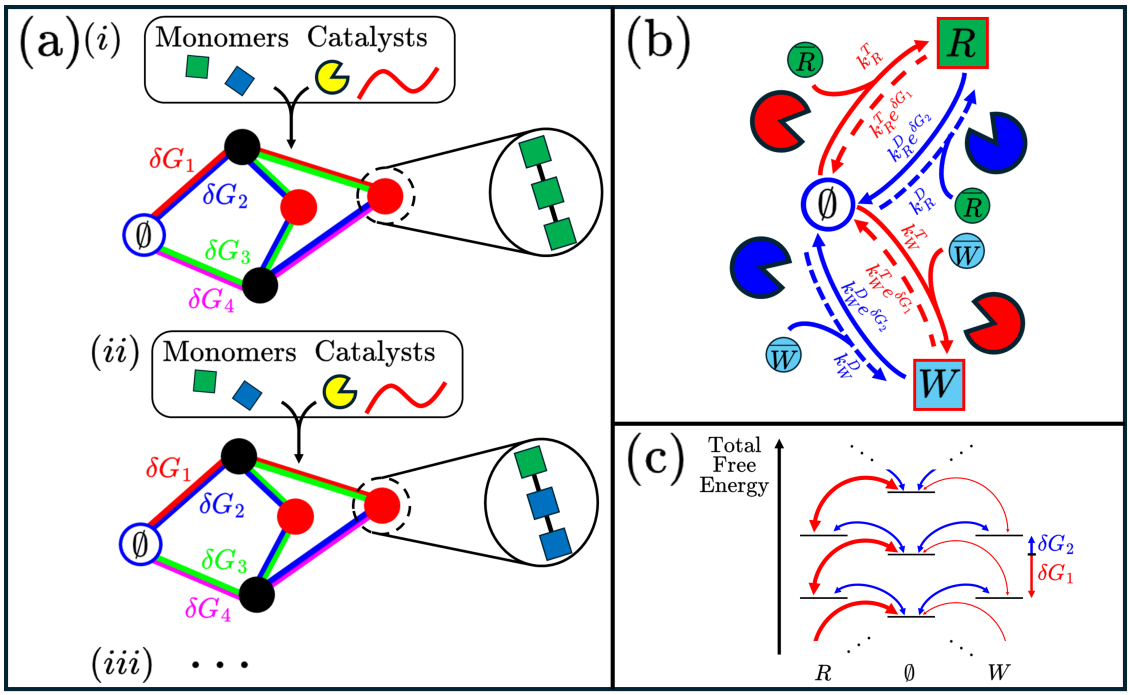


FIG. 1: Model systems in which products are selectively produced by catalysts from a pool of monomers. (a) A very simple example network illustrating the class of systems considered herein. Any one of a number of (polymer) products can be produced via equivalent pathways that rely on catalysed addition of input monomers. Both catalysts and monomers are assumed to be held at constant concentration by a chemostat, resulting in a linear CRN that can be represented as a graph. In (i), we illustrate a simple example of such a graph, including a null, polymer-free state; intermediate states represented by black circles and products represented by red circles. Each transition is reversible (arrowheads and rates are omitted for clarity). Each distinct pathway from the null state to the product is coloured a unique colour, and has a well-defined associated free-energy change δG. In (ii) we take an alternative product as the root of the graph; the network connecting it to the null state is topologically identical to (i). The equivalent pathways for each product have the same δG, but selective catalysts ensure that the rates associated with each pathway are product-dependent. (b) The simplest non-trivial system of this kind, in which monomers ($\bar{R}$ and $\bar{W}$) at equal concentration are converted to and from products R and W by two catalysts. R and W are connected to $\emptyset$ via topologically equivalent networks, involving a pathway via the red T catalyst and a pathway via the blue D catalyst. The free-energy change along each pathway is the same for each product; only the transition rates distinguish R and W. For simplicity, the short-lived catalyst-bound intermediates are not explicitly represented. We assume $\bar{R}$ and $\bar{W}$ are each chemostatted at the same concentration, which have been absorbed into the reaction rates. (c) Free energy landscape for the model in (b). The possibility of turning over fuel molecules introduces a ladder of R, W and null states, connected by the transitions involving the two catalysts (red and blue arrows, respectively). Overall δG from a given null state to given pair of R and W states is equal, but relative rates of downhill transitions (indicated by the thickness of the arrows) can be very different between R and W. In this case, the destructive (blue) catalyst is non-specific, with the same rates for R and W.

the same thermodynamic stability [2, 25, 29, 30, 32], consistent with the fact that catalysts cannot adjust the thermodynamic stability of the products that they select. Moreover, all products can be formed by a topologically equivalent set of processes (figure 1 (a)). By analogy with the processes of the central dogma, we shall loosely describe the catalysts that tend to generate products as "templates", but we do not make any particular assumptions about the detailed mechanisms.

Each pathway for creating or destroying a given product can incur a different free-energy change (figure 1 (a)), due to coupling with ancillary fuel molecules. These free-energy changes specify the equilibrium that would result in the long time limit if that pathway operated in isolation. We will show that this set of equilibrium free energies sets a lower bound on the entropy of the product distribution in a non-equilibrium steady state, for any fuel-coupled catalyst-induced kinetics. If the system is interpreted as an information channel between the input catalysts and the output products, this lower bound on entropy is equivalent to an upper bound on channel capacity [34, 35].

The bound on product distribution entropy also sets a bound on the accuracy with which a single product can

dominate the resultant ensemble. But surprisingly, the goals of producing a distribution maximally specific to a single chemical species and producing the distribution of products with minimal entropy are generally distinct. Moreover, contrary to our expectations, the optimal system that minimizes the entropy of the product distribution – pushing the distribution as far from equilibrium as possible – is not one in which desired products are systematically assembled on a template and degraded by another catalyst [31]. Rather, the optimal system is a pseudo-equilibrium one in which kinetic specificity ensures that each product is coupled overwhelmingly to only a single pathway, with each product's yield specified by the equilibrium properties of that pathway.

Exploring our bound, we observe that, for $M \to \infty$ possible products, it is possible to selectively produce a single product with perfect accuracy in steady state provided two pathways exist that are separated by ${}^{\Delta G}/_{\ln M} > k_{\mathrm{B}}T$ in their free-energy change, irrespective of other pathways. However, the probability of producing a single correct product as $M \to \infty$ is zero for ${}^{\Delta G}/_{\ln M} < k_{\mathrm{B}}T$. Further, the convergence to the infinite M behaviour is slow. Similarly, zero entropy product distributions can never be achieved with finite ΔG, but even low entropy requires $\Delta G > k_{\mathrm{B}}T(\ln M + \ln\ln M)$.

These results appear consistent with the notion that accurately maintaining a specific pool of products has a thermodynamic cost related to the information stored ($\ln M$ if a single template selects for a unique product) [2, 31, 32]. However, this simple picture is incomplete. Since the bound is saturated for pseudo-equilibrium systems in which each product couples overwhelmingly to a single pathway, there is no minimal entropy production per product molecule created in these steady states. Moreover, for $M \to \infty$, although a single product cannot be maintained with any accuracy for ${}^{\Delta G}/_{\ln M} < k_{\mathrm{B}}T$, it is possible to maintain a vanishingly small fraction of the total possible products with perfect accuracy for any $\Delta G \gg k_{\mathrm{B}}T$, even if $\Delta G \ll k_{\mathrm{B}}T \ln M$. Finally, the bounds only apply to steady states, and it is possible to produce distributions of arbitrary precision at finite time with no minimal ΔG.

In Section II, we formally introduce the class of models and techniques used to analyse them. In Section III, we motivate our analysis by considering the simplest possible case of a network of this kind, in which two products can both be produced via two pathways. We find that, even with full control over the forward fluxes of each reaction, the concentrations of each product are still constrained by the equilibrium properties of the individual pathways. In Section IV A, we then derive bounds on the distribution of products for a more general class of models that include multiple, sometimes overlapping creation and destruction pathways to an arbitrary number of products. In Section IV C, we discuss the physical interpretation of the bounds. Finally, in Sections V A and V B, we explore how systems are constrained by the bound in two biologically inspired examples. Throughout this article, we shall use Shannon entropy in units of nats, free energies in units of k_BT, and concentrations relative to a reference concentration. Additionally, we use the symbol "$\lesssim$" to mean "less than something which is approximately equal to" and similarly for "$\gtrsim$".

II. MODELS AND METHODS

A. Physical model

We now define the class of models considered, illustrated in figure 1 (a). Consider a set of input chemical species that may be assembled into a number of different possible products. For simplicity, and to focus on information propagation, we shall assume that all input species are held (chemostatted) at the same constant concentration by the cell or the environment. In these context of the central dogma of molecular biology, these inputs would be the nucleotides or amino acids. There are M possible different products that these inputs may form, corresponding to the different sequences of copolymers (*e.g.* RNAs, polypeptides). Finally, there are catalytic species (*e.g.* templates, polymerases, or nucleases) that facilitate the inter-conversion of inputs and products via intermediates. We assume that these catalysts are also chemostatted, and neglect any reaction involving more than one product and/or intermediate. We are interested in the set of products formed.

Such a model may be expressed as a deterministic chemical reaction network (CRN) [36, 37]. A deterministic CRN consists of a set of chemical species, a set of complexes (which are collections of species), and a set of reactions. In each reaction, a complex (the set of reactants) is converted into a different complex (the set of products) at some reaction rate, which is a function of the concentrations of species making up the reactants. We shall work with deterministic reversible CRNs obeying mass action kinetics. By reversible, we mean that, if a reaction from a complex $\mathcal{X}$ to a complex $\mathcal{Y}$ is allowed, then its reverse reaction from $\mathcal{Y}$ to $\mathcal{X}$ is also allowed. By mass action kinetics, we mean that the rate of a given reaction is proportional to the product of concentrations of species making up the reactants. Henceforth, we shall refer to the constant of proportionality as the "reaction rate". Since all reactions are reversible, we may use the principle of local detailed balance to define free-energy changes due to reactions. For a pair of reactions $\mathcal{X} \to \mathcal{Y}$, with reaction rate k^+, and its reverse $\mathcal{Y} \to \mathcal{X}$, with reaction rate k^-, the free-energy change due to this reaction is

$$\delta G_{\mathcal{X}\to\mathcal{Y}} = -\ln\left(\frac{k^+}{k^-}\right). \tag{1}$$

Given our assumptions about their chemostatting, the concentrations of input species and catalysts can be treated as parameters, rather than variables. Under

such assumptions, the CRNs are linear: the set of complexes contains only the null complex (containing no non-chemostatted species), and singletons (containing a single non-chemostatted species, *i.e.* an intermediate or a product). All reaction rates therefore depend linearly on the concentration of the non-chemostatted species. Transitions from the null complex correspond to the first association between inputs or inputs and catalysts to produce an intermediate.

We can draw a linear CRN as a graph (*e.g.* figure 1 (a)(i)), with nodes as species, edges as reactions and weights as reaction rates, absorbing the concentration of any chemostatted species involved into the reaction rates. We shall assume the graph produced is connected, as it must be if all species can be produced from the null state. This graph defines a set of self-avoiding walks (SAWs) from the null state to each product; each SAW together with its inverse define a "pathway" between the null state and the products.

We allow for the transitions to be coupled to implicit fuel molecules, meaning that different SAWs are associated with different δG. However, each edge must correspond to the turnover of a fixed number of fuel molecules; branching pathways allow for variable fuel consumption when coupled to a given catalyst. Each SAW from the null complex to a product therefore has a well-defined free-energy change.

We assume that products are all connected to the null complex by topologically equivalent graphs (figure 1 (a)). Such an assumption makes sense for a set of products corresponding to, for example, RNA molecules of a fixed length, which can all be grown via an equivalent set of steps. We further assume that all products are equally thermodynamically stable. This assumption is consistent with the fact that catalysts cannot thermodynamically favour one product over another [2, 25, 29, 30, 32], and allows us to focus on the question of how information is propagated from catalysts to products. Any internal thermodynamic bias towards certain products would not generally help to form an arbitrary product.

As a consequence, the equilibrium product state would contain a uniform distribution over products [2] and a narrower distribution over products can only be achieved via kinetic selection of a non-equilibrium steady state by the catalysts. Moreover, there exists a finite number of SAWs from the null complex to any given product, each with a free-energy change δG_i where i indexes the given SAW. The set of SAWs from the null complex to any product will be topologically equivalent and incur the same set of total free-energy changes (figure 1 (a)).

These constraints on free energy only apply to full paths from the null state to completed products. Intermediate states, involving partial, catalyst-bound products, can have sequence-specific free energies. Moreover, the transition rates between states can be different, subject to the overall constraint that, along a given complete SAW from null to product, the total free-energy change is the same as its equivalent for all other products. We will explore how relative yields of the products can be optimised by varying these strictly positive rates arbitrarily under this constraint.

Finally, we focus on deterministic CRNs here. However, since the CRNs are linear and connected, our results also apply to the expected concentrations at steady state of a stochastic realisation of the CRN [38].

B. Analytical methods

The steady state concentration of species X in a linear, deterministic CRN can be found directly from the graph. The concentration is given by the ratio of the sum of spanning trees of the graph rooted at the species X to the sum over spanning trees rooted at the null species. Explicitly,

$$c_X = \frac{\sum\limits_{T\in\mathcal{T}(X)}\prod\limits_{e\in T} k(e)}{\sum\limits_{T\in\mathcal{T}(\emptyset)}\prod\limits_{e\in T} k(e)}, \tag{2}$$

where $\mathcal{T}(X)$ is the set of spanning trees rooted at node X, $e \in T$ is an edge in graph T, and $k(e)$ means the weight of edge e. For a proof, see appendix B or [39] for a similar result.

These sums over spanning trees may be usefully rewritten as a sum over the SAWs from the null species to the given product species X [28, 40]. Having done so, the results are expressed in terms of the free-energy change along a given SAW [36] S,

$$\delta G_S = -\ln\left(\prod_{e\in S}\frac{k(e)}{k(\bar{e})}\right), \tag{3}$$

where $\bar{e}$ is the reverse reaction of edge e. Note that SAWs with positive δG_S correspond to pathways that tend to disassemble products.

Using graph theoretical methods, one may calculate the steady state concentrations for all of the product species. Having done so, and denoting the products by Z_i for $i = 1, \cdots M$, we define the product distribution as the probability that a randomly chosen product from the resultant ensemble is Z_i,

$$\mathbb{P}(Z_i) = p_i = \frac{c_{Z_i}}{\sum\limits_{j=1}^{M} c_{Z_j}} = \frac{c_{Z_i}}{c_T}, \tag{4}$$

where $c_T = \sum\limits_{j=1}^{M} c_{Z_i}$ is the total concentration. We define the (Shannon) entropy [35] of this distribution as

$$H[p_i] = -\sum_{i=1}^{M} p_i \ln p_i = \ln c_T - \frac{1}{c_T}\sum_{i=1}^{M} c_{Z_i}\ln c_{Z_i} \tag{5}$$

In appendix A, we show how under certain assumptions, the minimisation of this entropy maximises the channel

capacity of this system if treated as an information channel from catalysts to products. Informally, a lower entropy reduces the probability that the same product is sampled from the output of two different catalyst configurations.

C. Simulation of specific networks

Although calculations of the bounds (section IV A) is often straightforward, evaluation of the actual performance of a system realisation can be more challenging. For the examples presented in section V, steady state distributions are found by numerical solutions of the underlying ordinary differential equations (ODEs). The CRNs are linear, connected and contain only one stoichiometric compatibility class. Hence, there exists a single positive steady state to the ODEs induced by mass action kinetics [36]. Initially, all concentrations are set to zero, and the ODEs are simulated for a large time until no change to the distribution is observed. For section V B, to speed up simulation, we make use a of a result from [28] whereby we may coarse grain some sets of reactions without changing the steady state. The code used to simulate the ODEs is available from the zenodo repository.

III. FREE-ENERGY DIFFERENCES BETWEEN PATHWAYS BOUND ACCURACY FOR A SIMPLE NETWORK WITH TWO CATALYSTS

To build intuition, we first consider the simplest nontrivial system in which inputs and products are interconverted by catalysts (figure 1 (b,c)). The reaction network has two product species that we denote "right" (R) for the target species and "wrong" for the other species (W). These products are formed by catalytic activation of their inactive states $\bar{R}$ and $\bar{W}$, which are the input monomers, present at equal concentrations. Absorbing the concentrations of $\bar{R}$ and $\bar{W}$ and the catalysts into the rate constants, we may describe the system via the following CRN:

$$\begin{aligned} \emptyset \underset{k_R^T}{\overset{k_R^T}{\rightleftharpoons}} R, \quad \emptyset \underset{k_R^D e^{\delta G_f}}{\overset{k_R^D}{\rightleftharpoons}} R, \\ \emptyset \underset{k_W^T}{\overset{k_W^T}{\rightleftharpoons}} W, \quad \emptyset \underset{k_W^D e^{\delta G_f}}{\overset{k_W^D}{\rightleftharpoons}} W, \end{aligned} \tag{6}$$

where $\emptyset$ is the null species, and the effective reaction rates are written above/below the arrows. For each product there are two pathways from the null state leading to the product, one via a catalyst we call the template (reaction rates labelled by T in equation 6) and one via a destructive catalyst (reaction rates labelled by D in equation 6). For simplicity, we have assumed the catalyst-bound intermediates are short-lived, and do not explicitly represent them.

The pathways via the destructive catalyst are both biased with free energy $\delta G_2 = \delta G_f > 0$ against creating the product, relative to the pathways via the template ($\delta G_1 = 0$). A non-zero δG_f implies that the D (or T) pathway is coupled to the turnover of implicit molecular fuel, allowing the apparent disruption of detailed balance [33]. Each individual pathway, however, has a well-defined free-energy change associated with it; in the presence of only that pathway, the system would relax to an equilibrium defined by the free-energy change of that pathway.

The two products R and W can be produced by topologically equivalent pathways, by processes with the same overall free energy change. R and W are therefore both equally thermodynamically stable and a non-equilibrium excess of R over W is only possible if the template is kinetically selective for production of R over W. This setup is a minimal analog of a transcription-like system in which all products are thermodynamically viable, but those selected by the presence of their templates are preferentially produced.

In a conventional analysis of a system in which a template selectively produces a specific daughter molecule [9–29], it is typical to simply consider how selective the template is when producing the intended product. We can calculate the net flux of product, R, along pathway T in steady state as

$$J^T(R) = k_R^T(1 - c_R) = \frac{k_R^T k_R^D (e^{\delta G_f} - 1)}{k_R^T + k_R^D e^{\delta G_f}}, \tag{7}$$

where c_R is the concentration of R. Equivalent equations have the same form for $J^D(R)$, $J^T(W)$, and $J^D(R)$. At the steady state, $J^T(s) = -J^D(s)$ for $s = R, W$.

By varying the underlying rates at fixed δG_f, the flux ratio relating the production rate of R and W by T can be straightforwardly varied. A quick calculation shows that this ratio is

$$\frac{J^T(R)}{J^T(W)} = \left(\frac{k_R^T k_R^D}{k_R^T + k_R^D e^{\delta G_f}}\right) \Big/ \left(\frac{k_W^T k_W^D}{k_W^T + k_W^D e^{\delta G_f}}\right). \tag{8}$$

This ratio can be made arbitrarily large for any δG_f, simply by picking appropriate rates $k_R^T \gg k_W^T$, forcing the template to produce output R at a rate arbitrarily higher than the rate it produces output W. The catalyst can therefore be arbitrarily kinetically selective for a given δG_f along each pathway; kinetic selectivity of catalysts is not – in principle – constrained by thermodynamics.

We might, therefore, expect to be able to produce an output distribution that contains an arbitrarily higher concentration of R than W by varying the reaction rates at fixed δG_f. However, this is not the case. The steady-state concentrations of R and W are

$$c_R = \frac{k_R^T + k_R^D}{k_R^T + k_R^D e^{\delta G_f}}, c_W = \frac{k_W^T + k_W^D}{k_W^T + k_W^D e^{\delta G_f}}. \tag{9}$$

Both these concentrations are bounded by $e^{-\delta G_f} \leq c_{R/W} \leq 1$. Hence, we cannot achieve an

arbitrary ratio of concentrations, c_R/c_W. Moreover, since ratios $J^T(R)/J^T(W)$ and c_R/c_W can be set independently, an arbitrarily high ratio of fluxes does not necessarily correspond to a high ratio of concentrations. For example, if the rates obey $k_R^T \gg k_R^D, k_W^D \gg k_W^T, k_R^D \gg k_W^T$ (figure 1 (c)) then $J^T(R)/J^T(W)$ will be large, and c_R/c_W will be close to $e^{\delta G_f}$. However, if instead $k_R^T \gg k_R^D, k_W^D \gg k_W^T, k_W^T \gg k_R^D$, then c_R/c_W will still be close to $e^{\delta G_f}$, but $J^T(R)/J^T(W)$ will be small.

The optimal distribution giving the highest proportion of the right product R is found when $k_W^T \to 0$ and $k_R^D \to 0$. In this case, R is overwhelmingly produced and destroyed via the template. For R, it is as if the yield is governed by the equilibrium of a single red transition in figure 1 (c). W, by contrast, is overwhelmingly produced and destroyed via D; its yield is equivalent to the equilibrium of a single blue transition in figure 1 (c). This "pseudo-equilibrium" limit gives a concentration of $c_R = 1$ for R and a concentration of $c_W = e^{-\delta G_f}$ for W. We note that although the individual products appear to equilibrate, the product ensemble remains far from equilibrium, because different product species are coupled to different equilibria.

In conclusion, even with arbitrary relative fluxes of different products along different pathways, and a far-from-equilibrium distribution of products, the relative steady-state concentrations of products are bound by the differences in the equilibrium properties of individual pathways, and the ideal behaviour involves vanishing cyclic flux around the template/destroyer cycle. In the remainder of this paper, we shall generalise this result and explore its consequences.

IV. GENERAL BOUNDS ON THE PROPERTIES OF PRODUCT DISTRIBUTIONS

A. Derivation of the bounds

We shall now present the main mathematical results of this work: an upper bound on the maximal probability of a single product $p_{\max}$, and a lower bound on the entropy $H[p_i]$ of the product distribution. To do so, we first note that for the class of linearised CRNs considered here, the steady-state concentration of any species is bound by the set of free-energy changes along the pathways that create or destroy it. Generalizing the result in Section III, the steady state concentration is upper-bounded by the equilibrium implied by the single pathway with the most negative free-energy change of species formation, and lower-bounded by the equilibrium implied by the pathway with the most positive free-energy change of species formation. These bounds are achieved when the rates along all other pathways tend to zero.

Mathematically, we define $\delta G_L^{Z_i} = \max_{SAWs,S} \delta G_S^{Z_i}$ and $\delta G_U^{Z_i} = \min_{SAWs,S} \delta G_S^{Z_i}$ as the maximal and minimal free energy changes along self-avoiding walks that lead to the creation of Z_i from the null state, respectively. Then

$$e^{-\delta G_L^{Z_i}} \leq c_{Z_i} \leq e^{-\delta G_U^{Z_i}}. \tag{10}$$

The proof for this statement is given in Appendix C, although similar bounds are shown in [39–42].

Our systems have the particular property that the concentrations of all product species have the same upper and lower bounds $(e^{-\delta G_U}, e^{-\delta G_L})$. As a result of this crucial fact, properties of the distribution defined by eq. 4 are constrained. Specifically, the maximal probability of a single product has a non-trivial upper bound, while the entropy has a non-zero lower bound. We now calculate the optimal distribution for maximising one desired product and the distribution for minimising total product distribution entropy, showing that they are different.

When maximising the probability of a single product species, hereafter referred to as *specificity maximisation*, the probability distribution will be one product with probability $p_{\max}$, and all others with p_{low}, with

$$\begin{aligned} p_{\max} &= \left(1 + (M-1)e^{-\Delta G}\right)^{-1}, \\ p_{\text{low}} &= e^{-\Delta G}\left(1 + (M-1)e^{-\Delta G}\right)^{-1}, \end{aligned} \tag{11}$$

where $\Delta G = \delta G_L - \delta G_U$ and M is the number of possible product species.

Alternatively, we can consider *entropy minimisation*, or finding the minimal entropy of a distribution under the constraint that all species are restricted by $e^{-\delta G_L} \leq c_i \leq e^{-\delta G_U}$. We obtain $H[p_i] \geq H_{\min}$, with

$$H_{\min} = \frac{(M-m)\Delta G e^{-\Delta G}}{m + (M-m)e^{-\Delta G}} + \ln\left(m + (M-m)e^{-\Delta G}\right). \tag{12}$$

$H_{\min}$ arises from a distribution in which m species are at the upper bound concentration, $\exp(-\delta G_U)$, and $M-m$ species are at the lower bound, $\exp(-\delta G_L)$. The m that minimises entropy is given by either:

$$\left\lceil \frac{e^{-\Delta G}\left[\Delta G - \left(1 - e^{-\Delta G}\right)\right]}{\left(1 - e^{-\Delta G}\right)^2} M \right\rceil$$
or
$$\left\lceil \frac{e^{-\Delta G}\left[\Delta G - \left(1 - e^{-\Delta G}\right)\right]}{\left(1 - e^{-\Delta G}\right)^2} M \right\rceil - 1, \tag{13}$$

whichever gives the lower H. We use the ceiling function because m is integer; we are required to check two values on either side of the minimum of the continuous function. A reasonable approximation to the optimal value of m is given by

$$m_{\min} = \max\left(\frac{e^{-\Delta G}\left[\Delta G - \left(1 - e^{-\Delta G}\right)\right]}{\left(1 - e^{-\Delta G}\right)^2} M, 1 \right), \tag{14}$$

since for $m_{\min} > 1$, the difference between using the integer value of $m_{\min}$ (eq. 13) and the continuous value (eq. 14) is small. In practice, using equation 14 produces

a very slightly looser bound on $H[p_i]$. Further, using eq. 14, for $m_{\min} > 1$, we may simplify:

$$H_{\min} = \ln M - \Delta G \left(1 + \frac{e^{-\Delta G}}{1 - e^{-\Delta G}}\right) + \ln\left(\frac{\Delta G}{1 + e^{-\Delta G}}\right) + 1. \tag{15}$$

The proofs for these results are given in appendix D. Note that in the context of copolymerisation, $\ln M$ is proportional to the length of polymer necessary to produce M different possible product sequences.

B. The bounds in the presence of kinetic proofreading

Surprisingly, internal cycles in the reaction network, a requirement for celebrated kinetic proofreading motifs [4–6], do not directly feature in the bounds derived, since SAWs cannot contain cycles by definition. Adding a proofreading loop to a process may affect ΔG, since it may provide a new pathway with a maximal or minimal free-energy change, but the SAWs identified would always be loop-free and correspond to a fixed free-energy change for product formation. The possibility of repeatedly undergoing a single dissipative cycle in an actual dynamic trajectory, consuming an arbitrary amount of molecular fuel, does not translate into an arbitrary δG along a pathway. However, the existence of loops within the network may still help to achieve a better product distribution than otherwise if transition rates are constrained beyond the free-energy changes along SAWs being fixed.

C. Physical significance of the bounds

We split the analysis into three main parts. In the first, we discuss the difference between specificity maximisation and entropy minimisation. We look at the properties of both of these distributions in detail. In the second we compare our result to the previous arguments from Bennett on the minimal cost of systems that create product polymers via templates and destroy them via other pathways [31]. Finally, we discuss whether and how systems can, even in principle, reach the bounds we find.

1. *Entropy minimisation versus specificity maximisation*

One might naively expect that the distribution of product concentrations that minimises the entropy would be the same as the distribution that optimises the probability of a single product. However, this is not the case in general. The concentration of each species is constrained between $e^{-\delta G_L}$ and $e^{-\delta G_U}$. The lower concentration cannot be arbitrarily close to zero, and the higher concentration cannot be arbitrarily high. Since having multiple products at the high concentration increases the total concentration of products, suppressing the relative probabilities of sampling products at the low concentration, the probability distribution may be sharper for systems with multiple products at a higher concentration. In this subsection, we first discuss when specificity maximisation and entropy minimisation are equivalent; then proceed to analyse specificity maximisation and finally entropy minimisation.

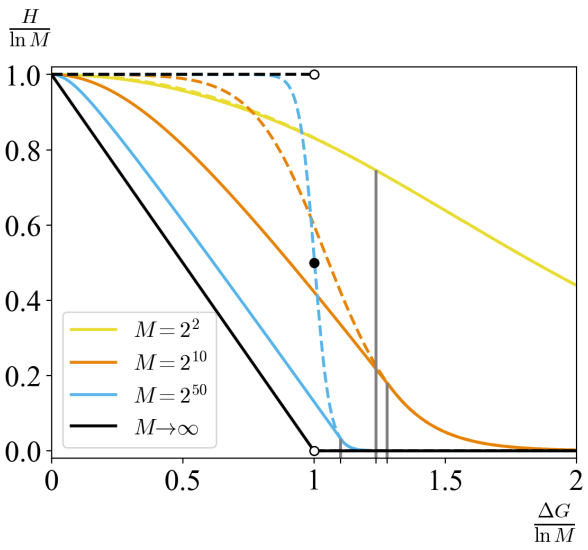


FIG. 2: Entropy minimisation is not equivalent to specificity maximisation when $\Delta G \lesssim \ln M + \ln\ln M$. We plot $H_{\min}$ (solid lines) and the entropy of the distribution for specificity maximization (dotted lines) as a function of ΔG for various M. We scale both quantities by $\ln M$. The dashed line is discontinuous at $\Delta G = \ln M$ for $M \to \infty$. The vertical (gray) lines show $\Delta G = \ln M + \ln\ln M$. For $\Delta G \lesssim \ln M + \ln\ln M$, the minimum entropy distribution has a significantly lower entropy than the maximum specificity distribution. For $\Delta G \gtrsim \ln M + \ln\ln M$, the dotted and solid lines merge, as the minimum entropy distribution is the maximal specificity distribution. As M increases, the solid lines tend to the solid black line, albeit very slowly.

If $m = 1$ minimises eq. 12, entropy minimisation and specificity maximisation are equivalent. The lowest entropy state then contains a single product at a high concentration, and all others at a low concentration. We find $m_{\min} = 1$ is optimal for $\Delta G > \ln M + \ln\ln M + \mathcal{O}\left(\frac{\ln\ln M}{\ln M}\right)$ (figure 2). For ΔG smaller than this value, the entropy of the distribution that maximises specificity differs quite drastically from that minimising the entropy. Surprisingly, the distribution for maximising specificity has an entropy consistent with an unbiased distribution ($H = \ln M$) for $\Delta G/\ln M \leq 1$ in the limit $M \to \infty$. In other words, the concentration of the desired product tends to zero relative to the total concentration of products.

To explore this observation further, we consider the

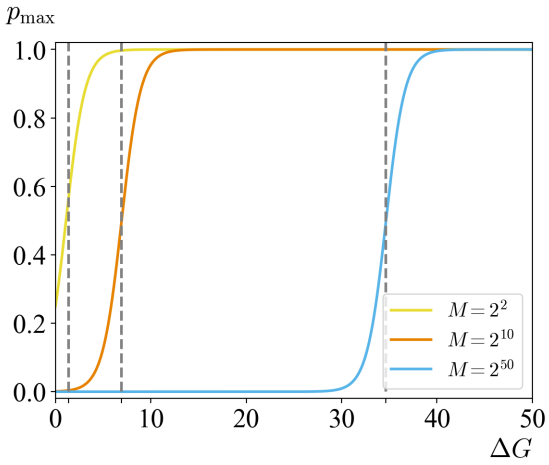
maximal product probability $p_{\max}$. The probability $p_{\max}$ is a sigmoidal function centred on $\Delta G = \ln M$, with $p_{\max} = M^{-1}$ at $\Delta G = 0$ and $p_{\max} \to 1$ as $\Delta G \to \infty$. Interestingly, the sigmoidal function does not approach a step function here, but has fixed width around $\ln M$ (figure 3a). Instead $p_{\max}$ has the exact form $p_{\max} = (1+e^{-(\Delta G - \ln M)})^{-1}$ as $M \to \infty$. Thus, if we were to plot $p_{\max}$ as a function of $\Delta G - \ln M$, the curves for different, but large enough, M would collapse onto each other. Therefore, to achieve a $p_{\max} \approx 1 - e^{-A}$ for small e^{-A}, one needs a $\Delta G \approx \ln M + A$. If we instead scale by system size, we find that $p_{\max}$ tends to a step function when $M \to \infty$ (figure 3b), reiterating that the relative yield of a single product is necessarily zero below ${}^{\Delta G}/_{\ln M} = 1$ as $M \to \infty$. Although the intended product can be far more probable than any single alternative, the large number of possible alternatives dominates.

To put the bound on $p_{\max}$ into context, we consider the implications for the charging of a single tRNA with the correct amino acid. Here, there are approximately $M = 400$ different combinations of codon and amino acid, having grouped all codons that correspond to the same amino acid into one. Using $M = 400$, we can calculate the bound for the specificity with which an ensemble dominated by a single charged tRNA could be maintained in steady state. We see that for an error rate of 10^{-A}, *i.e.* $p_{\max} = 1 - 10^{-A}$, one needs a free-energy difference in the pathways connecting inputs and products of approximately $(2.3A+6)k_BT$. Thus, for an error rate of 10^{-5}, one would need a free-energy difference of around $17k_BT$ or 1 ATP[43] at $37°C$, substantially higher than $kT \ln M = k_BT \ln 400 \approx 6k_BT$, the value implied by the large M limit.

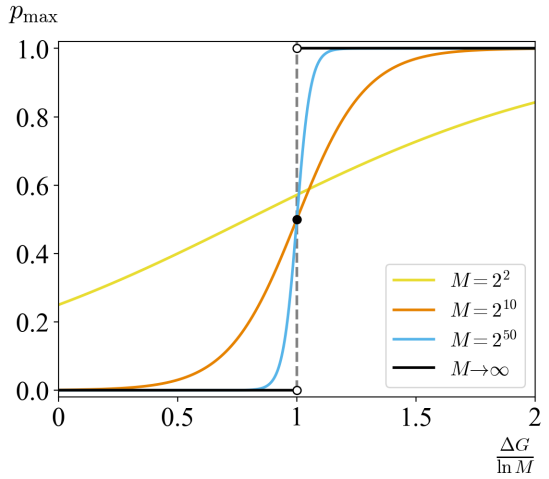
For entropy minimisation, we find that the limits $\Delta G = 0$ and $\Delta G \to \infty$ are identical to those found by specificity maximisation (figure 2). At $\Delta G = 0$, all pathways carry the same free-energy change and so the system is a true equilibrium one and no specificity is possible, $H_{\min} = \ln M$. For $\Delta G \to \infty$, $H_{\min} \to 0$ for all M and a single product dominates. As can be seen from figure 2, the approach to the $M \to \infty$ limit is very slow. Even for $M = 2^{50}$, there is a significant discrepancy between $H_{\min}$ and the minimal entropy for the $M \to \infty$ limit. This difference, when normalised by $\ln M$, is approximately ${}^{\ln \ln M}/_{\ln M}$ (which is ≈ 0.1 for $M = 2^{50}$).

We now consider the properties of the entropy-minimizing distribution for $m_{\min} > 1$. Here, $m_{\min}$ species are at the same high concentration, and the remaining species are at a low concentration. The fraction $m_{\min}/M$ given by eq. 14 is independent of M, and monotonically decreases to 0 with increasing ΔG. Thus, if we scale $\Delta G \to \infty$ at fixed ${}^{\Delta G}/_{\ln M}$, the fraction of products with high yields will tend to zero even for $\Delta G \ll \ln M$, well inside the region where specificity for a single product is impossible (figure 4a).

Furthermore, the probability of a product randomly chosen from this distribution being one of the $m_{\min}$ with a high concentration is given by $p_{\text{high}} = 1 - \frac{1}{\Delta G} + \frac{1}{e^{\Delta G}-1}$.



(a)

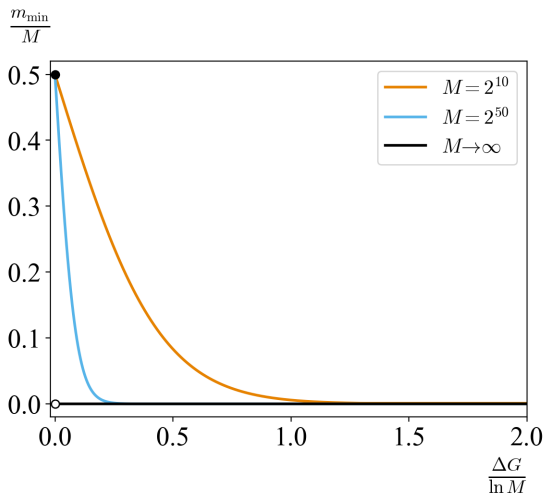


(b)

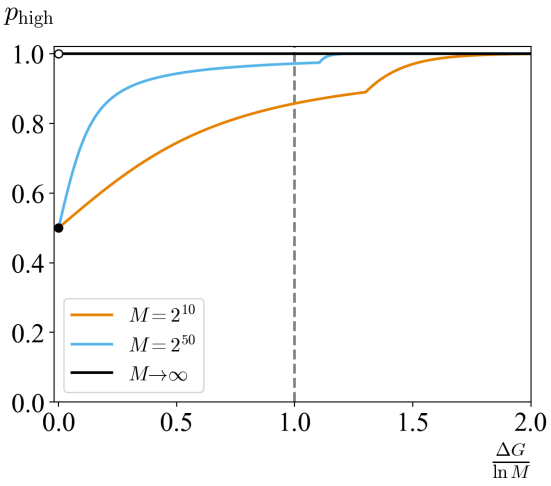
FIG. 3: Specificity maximisation results in a $p_{\max}$ that is a sigmoidal function of ΔG centred at $\Delta G = \ln M$. $p_{\max}$ for a specificity-maximised system (a) as a function of ΔG and (b) as a function of scaled free energy ${}^{\Delta G}/_{\ln M}$. For (a), we see that the shape of the curves quickly approaches the same functional form, but shifted by $\ln M$. For (b), as M increases, $p_{\max}$ tends to a step function at ${}^{\Delta G}/_{\ln M} = 1$.

This quantity is also independent of M and increases monotonically with ΔG to 1 (figure 4b). Thus, as $M \to \infty$ at fixed ${}^{\Delta G}/_{\ln M}$, a vanishingly small proportion of the total number of products can contain an overwhelming majority of the probability, even for $\Delta G \ll \ln M$ and $m_{\min} > 1$. The black lines in figure 4 illustrate this surprising phenomenon in the limit of $M \to \infty$.

Turning now to ${}^{\Delta G}/_{\ln M} > 1$, we consider how $H_{\min}$ behaves for large ΔG and large, but finite, M. Below $\Delta G \approx \ln M + \ln \ln M$, from eq. 15, $H_{\min} \approx \ln M - \Delta G + \ln \Delta G + 1$, which is approximately linear in ΔG (figure 5a). Above $\Delta G \approx \ln M + \ln \ln M$, $H_{\min}$ decreases



(a)



(b)

FIG. 4: The fraction $m_{\min}/M$ of species at the high concentration bound in the entropy minimizing distribution tends to zero with scaled free energy $\Delta G/\ln M$, while the probability of picking one of these high concentration species from an ensemble, p_{high}, tends to 1. (a) The fraction of products at high concentration in the entropy-minimizing distribution against $\Delta G/\ln M$ for two values of M and the limit as $M \to \infty$. We approximate this fraction as continuous using eq. 14. (b) The total probability of products at the high concentration. Notice how, even below $\Delta G/\ln M = 1$, p_{high} can be arbitrarily close to 1 and $m_{\min}/M$ arbitrarily close to 0 for large enough M.

exponentially as $H_{\min} \approx \exp(-(\Delta G - \ln M - \ln\ln M))$. Further, at $\Delta G \approx \ln M + \ln\ln M$, $H_{\min} \approx 1$. Since the maximum value of H is $\ln M$, this value for ΔG gives the approximate scale for $H_{\min}$ to be considered small.

In figure 5b, we show how the limits on the unnormalised entropy $H_{\min}$ vary with $\Delta G/\ln M$. Here, for sufficiently large $\Delta G \gtrsim 1.3 \ln M$, increasing M leads to a nominal decrease in $H_{\min}$, not just a decrease in $H_{\min}/\ln M$, as might be expected. Similarly for specificity maximisation (figure 3a), for sufficiently large $\Delta G \gtrsim 1.1 \ln M$, $p_{\max}$ increases with increasing M. Longer polymers therefore have much more scope within the bounds derived to use a given ΔG per monomer to suppress the absolute entropy of the resultant product distribution.

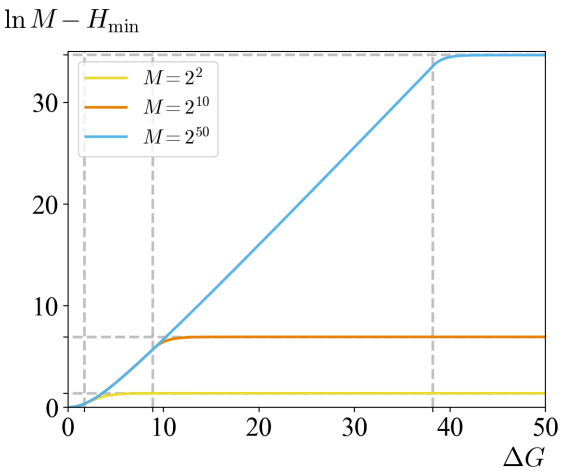
2. *Relationship of the bounds to previously postulated costs*

In his seminal paper [31], Bennett briefly discussed the production of a single product RNA copolymer using a specific template, and its subsequent destruction via a non-specific destructive enzyme. Without considering a specific model, he stated that since the enzyme is non-specific, one requires four times the phosphate concentration driving the non-specific destructive enzyme than one would if it was a sequence-specific destructive enzyme. This difference would prevent the non-specific destructive enzyme from creating non-target, random sequences, instead encouraging it to indiscriminately destroy everything, including the target sequence. He therefore states that a cycle of create via template followed by destruction via destructive catalyst must use at least $k_B T \ln 4$ per nucleotide of free energy, even in steady state, corresponding to $\Delta G/\ln M > 1$ in our formalism.

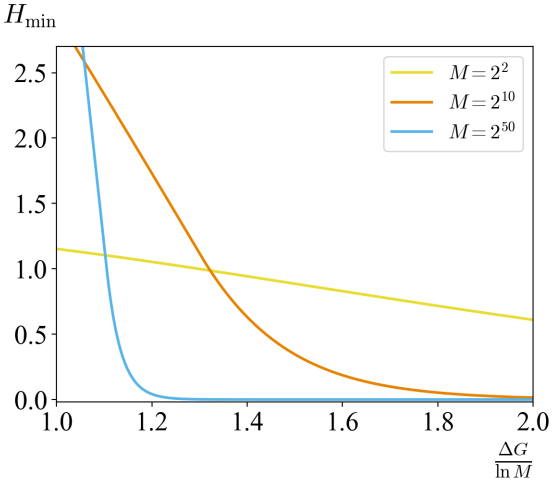
Our bounds update this understanding in several ways. Firstly, we point out that although two pathways with $\Delta G/\ln M > 1$ are required to generate a product distribution consisting of a single sequence in the steady state, the optimal system in terms of maintaining a specific product pool would not typically push polymers around a cycle, producing ΔG of entropy each time a polymer was created then destroyed. Instead, although the distribution over products as a whole is far from equilibrium, each product is effectively in its own kinetically-selected pseudo-equilibrium state. Each product is produced and destroyed by inverse pathways coupled to the same type of catalyst. Such a system has vanishing entropy production per creation event in the steady state.

Secondly, as shown through figures 2 and 3, while $\Delta G/\ln M > 1$ does serve as a necessary condition for perfect specificity in the limit of infinite polymer length, the approach to this limit is slow. A finite size correction on the order of $\ln L/L$ (where $L \propto \ln M$ would be the length of a polymer product) provides the minimum free-energy difference above $\Delta G = \ln M$ for high specificity to be achievable.

Thirdly, the bounds are fundamentally a property of distributions of products and monomers, rather than the specificity of templates [29]. Although the kinetic selectivity of the template is important in determining how close a system can get to the bound (see section V), it doesn't set the bound. This observation resolves the paradox that there are thermodynamic constraints on the templated ensemble, even though there are no thermody-



(a)



(b)

FIG. 5: $H_{\min}$ decreases roughly linearly with ΔG until $\Delta G \approx \ln M + \ln\ln M$, after which it decreases exponentially. We plot (a) the entropy drop $\ln M - H_{\min}$ against ΔG for various M, and (b) the entropy $H_{\min}$ as a function of the free-energy difference scaled by system size $\Delta G/\ln M$. In (a), The entropy drop is independent of M so long as $m_{\min} > 1$ (eq. 14). For large enough ΔG, $\ln M - H_{\min} \approx \Delta G - \ln \Delta G - 1$ (eq. 15), which is roughly linear, until $\Delta G \approx \ln M + \ln\ln M$ (vertical lines) whereupon the entropy drop ceases to be linear and saturates exponentially to $H_{\min} = 0$ (horizontal lines). In (b), above certain values of ΔG, the unnormalised entropy $H_{\min}$ is smaller for larger M at fixed $\Delta G/\ln M$.

namic constraints on catalytic specificity.

That the bounds are unrelated to catalytic specificity is particularly emphasised by considering the region $\Delta G/\ln M < 1$. In this region, for large M, a fraction of the possible products $m_{\min}/M$ dominates the ensemble. In principle, such an ensemble could follow from $m_{\min}/M$ distinct templates each catalysing formation of a single product sequence with high accuracy, even for $\Delta G/\ln M < 1$. Intriguingly, for $M \to \infty$, a vanishingly small fraction $m_{\min}/M$ of the possible products can constitute the full ensemble, in the absence of any other products, even for $\Delta G \ll \ln M$.

Fourthly, the limits on entropy and specificity with ΔG hold not only in the case of a non-specific destructive ecatalyst but also for a specific one, or indeed any complex set of pathways between monomers and products.

Finally, although our work mainly focuses on the issue of creating a steady state distribution of high specificity, we note by simple example that away from steady state, no free-energy differences are required to achieve arbitrary specificity. Consider a simple system with two products as follows:

$$\emptyset \underset{k_R}{\overset{k_R}{\rightleftharpoons}} R, \ \emptyset \underset{k_W}{\overset{k_W}{\rightleftharpoons}} W. \tag{16}$$

This CRN is a simplified version of the system considered in section III, with only a single catalyst and thus $\Delta G = 0$. The steady state bound on this distribution would give an unbiased equilibrium of $p_R = p_W = 1/2$. However, the concentrations of R and W can be very different at short times. Starting with initial conditions $c_R(0) = c_W(0) = 0$,

$$\lim_{t\to 0} p_R(t) = \lim_{t\to 0} \frac{c_R(t)}{c_R(t) + c_W(t)} = \frac{k_R}{k_R + k_W}, \tag{17}$$

which is only bounded by 0,1. Thus, the entropy can temporarily be below our steady-state bound and we can achieve arbitrarily high specificity for any set of free-energy differences between pathways.

3. *The achievability of the bounds*

The bounds we have derived are not always achievable, even if we allow a system to have arbitrary reaction rates consistent with the overall ΔG of each path. Whether or not the bounds are achievable depends on the topology of the CRN. To achieve the bounds, it is essential that it is possible to make the pathways corresponding to δG_U for m high-yield products and δG_L for $M - m$ low-yield products fast, while keeping all other pathways slow. Given that pathways to a product will typically share edges with pathways to other products, this independent scaling is not always possible.

A pleasing aspect of our result is that, although finding full steady-state concentration requires knowledge of all of the SAWs from $\emptyset$ to the product, identifying the bounds only requires knowledge of the pathways with the highest and lowest free-energy-change. These free energy changes can often be intuited or proven without knowing all of the pathways. Further, it is then often possible to check whether or not the set of pathways identified is indeed compatible with reaching the bound.

V. EXAMPLES

Having explored the properties of the bound itself, we consider two example systems to see how/whether the bound is reached as the rates along different edges are extremised.

A. Example 1: Templated polymerisation with a destructive catalyst

Refs [25, 28] studied a system in which a polymer is grown on a template. Monomers of either the "right" or the "wrong" type are added one at a time to a polymer in contact with a template, while the product polymer continually unbinds from the template from behind its leading edge. We now extend the system to include final dissociation from the template for complete polymers, and also a "destructive template". This destructive template participates in identical reactions to the template, except that polymerisation is driven backwards due to consumption of fuel molecules, meaning products tend to be destroyed rather than grown.

A full CRN for this model is shown in appendix E; we illustrate the linearised network (assuming monomers and catalysts are coupled to chemostats) for the case of dimerisation in figure 6. The thermodynamics of the model are characterised by the standard polymerisation free energy of the monomers ($-\delta G_{\mathrm{pol}}$); the standard free-energy change of binding to the template for right and wrong monomers ($-\delta G_R$ and $-\delta G_W$), and the free energy of fuel turnover $-\delta G_f$. We assume that both right and wrong monomers are held at concentration c.

For the model depicted in figure 6, we can calculate the free-energy change for different paths to each product state. For example, the paths from the null state to RR are:

$$\begin{aligned}
&1)\ \emptyset \to TR \to TRR \to RR,\\
&2)\ \emptyset \to TR \to TRW \to RW \to DRW\\
&\qquad\quad \to DR \to DRR \to RR,\\
&3)\ \emptyset \to DR \to DRW \to RW \to TRW\\
&\qquad\quad \to TR \to TRR \to RR,\\
&4)\ \emptyset \to DR \to DRR \to RR.
\end{aligned} \tag{18}$$

These incur free-energy changes:

$$\begin{aligned}
&1)\ -(\delta G_{\mathrm{pol}} + 2\ln c),\\
&2)\ -(\delta G_{\mathrm{pol}} + 2\ln c),\\
&3)\ -(\delta G_{\mathrm{pol}} + 2\ln c) + \delta G_f,\\
&4)\ -(\delta G_{\mathrm{pol}} + 2\ln c) + 2\delta G_f.
\end{aligned} \tag{19}$$

Each pathway contains the terms $-(\delta G_{\mathrm{pol}} + 2\ln c)$, corresponding to the standard free-energy change of product formation without any fuel turnover. Since our bounds only rely on the differences in free energies between pathways, we may drop these contributions, and the fuel free energy δG_f alone determines the bounds. For polymers

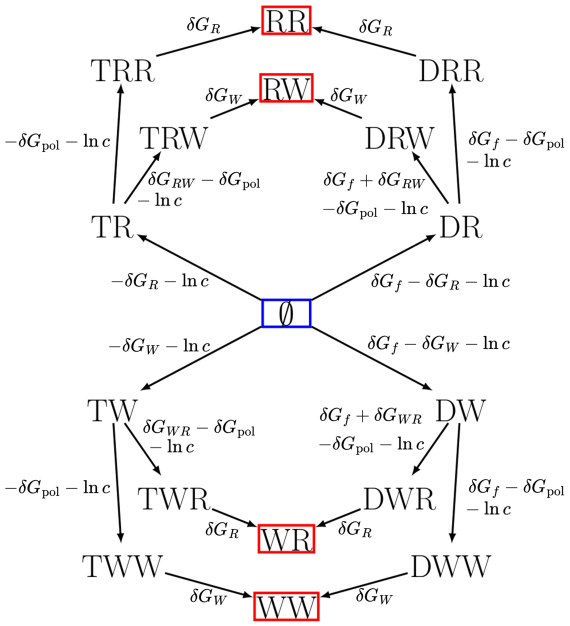


FIG. 6: Linearised CRN for a system in which dimers are grown/destroyed via a template "T" and by a destructive template "D". The system starts in the null state (blue) and a right or wrong ("R" and "W") monomer attach to either the template "T" or destructive enzyme "D". A second monomer can attach and polymerize with the first, yielding to a dimer that then detaches, giving the four red output states. Each arrow represents a reversible reaction, with the free-energy change in direction of the arrow indicated. For brevity, $\delta G_{RW} = -\delta G_{WR} = \delta G_R - \delta G_W$. To reach each red product node, there are four self-avoiding walks from the blue null state.

of length L, the equivalent standard free energy of product formation is $(L-1)\delta G_{\mathrm{pol}} + L\ln c$, and it too may be dropped for consideration of the bounds.

Our bounds are achieved when rates of the path with the most negative free-energy change are maximised for the desired product(s), and rates of the path with the most positive free-energy change are maximised for all other products. For the $L = 2$ case, for example, we could therefore maximise the rate of path 1 for RR, and the equivalents of path 4 for other products.

In fact, for this system, the entropy bound is formally achievable for arbitrary L. We can split the edges of the graph into two sets; one in which the rates are ~ 1, and one in which the rates are $\sim k$. In figure 7, we show $H[p_i]$ as $k \to 0$ for a particular choice of these sets, in which all the reactions leading directly to the fully correct sequence being created/destroyed on the template are set to be fast (not proportional to k), as are all the reactions leading directly to the creation/destruction of other sequences on the destroyer. The system saturates the entropy bound found in section IV A as $k \to 0$. Note that for the value of δG_f used, the minimal entropy and maximal specificity

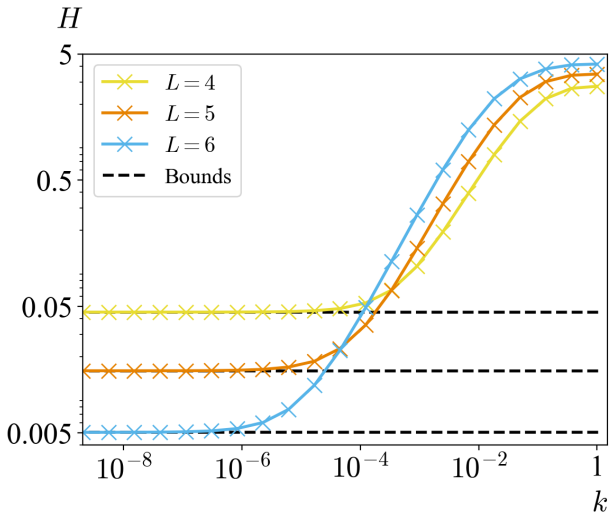


FIG. 7: The entropy bound may be saturated by model systems in the limit that some reaction rates are much smaller than others. We plot the entropy, H, of the product distribution as a function of the slow reaction rates, k, for different template lengths, L, for the simple production and destruction model introduced in section V A. The data is obtained for a fixed $\delta G_f = 2$, $\delta G_{\text{pol}} = 0$, $\delta G_R = 2$, $\delta G_W = -2$ and $c = 1$. Note that longer templates reach a lower entropy bound for a fixed fuel turnover per unit length, δG_f.

distributions are the same.

There are many possible ways to chose sets of edges that can saturate the bound in the limit $k \to 0$. Here, we have chosen a set of rates that specifically highlights a full pathway to each product for illustrative purposes. One might also wish to choose a minimum set of rates to have rate k while still saturating the bound in the limit $k \to 0$. For example, letting the slow reactions be $DR^L \to R^L$, where R^L means L copies of R, and $TX_1...X_L \to X_1..X_L$, excluding $X_1 \cdots X_L$ all being R (appendix E), will still saturate the bound in the limit $k \to 0$. Further, we note that it is possible to saturate the bound with a non-specific destructive catalyst, where the reaction rates are independent of the polymer sequence. In the examples we have identified, such a network requires at least three rate scales $\sim 1, k, k^2$.

We stress that although the bound is formally attainable in this system, doing so relies on the ability to manipulate rate constants arbitrarily, subject to thermodynamic constraints. In a more realistic model of a templating system, constraints on relative rates may also be relevant; these constraints may stop the system reaching the bounds on accuracy or product entropy.

B. Example 2: Hopfield-like kinetic proofreading with a destructive catalyst

To illustrate the application of the bound to a more complex network, and to demonstrate the possibility of non-trivial pathways defining the bound, we consider an extension to the previous model, wherein the template also performs kinetic proofreading. First suggested by Hopfield [4] and Ninio [5] and widely studied [6, 44], kinetic proofreading is a mechanism by which a system can increase the specificity of a process by expending extra free energy through fuel consuming cycles. These cycles give an extra opportunity to reject the "wrong" monomers due to their shorter binding lifetime.

The full chemical reaction network for a proofreading template of arbitrary length is given in appendix F. Once again, we linearise the system by assuming that monomers and catalysts are coupled to chemostats. In figure 8, we show part of the linearised CRN graph; this fragment should be inserted into figure 6 in place of the pathway ($\emptyset \to TR \to TRR \to RR$), with similar modifications to all other template-based pathways to RW, WR, and WW. Proofreading adds complexity to the graph in the form of additional loops, and we now explicitly consider a polymerization step independently from the binding to the template.

Monomers are now present in inactive (starred) and active forms, with $-\delta G_{\text{act}}$, representing the free-energy change of activation. We assume that each non-activated monomer R^*, W^* is chemostatted at the same concentration as each non-activated monomer R, W, c. Dropping this assumption would only cause a shift to δG_{act}. As a result, δG_f and δG_{act} control the concentration bounds. We assume δG_f, $\delta G_{\text{act}} \geq 0$.

The concentration bounds of this system, determined by the SAWs with maximal and minimal free-energy changes, depend on the values of both the aforementioned free-energy parameters, as well as whether L is odd or even. The bounds are set by

$$\delta G_U = \left\{ \begin{array}{c} \frac{L^2}{4} + \frac{L}{2} \\ \frac{(L+1)^2}{4} \end{array} \right\} \delta G_{\text{act}}, \tag{20}$$

and

$$\delta G_L = -\left\{ \begin{array}{c} \frac{L^2}{4} \\ \frac{L^2-1}{4} \end{array} \right\} \delta G_{\text{act}} - L\delta G_f \tag{21}$$

where the top value in each brace is for even L and the bottom value for odd L. Unlike the simple system in section V A, the SAWs that exhibit δG_U and δG_L are not the intuitively simple pathways that go via the template and the destroyer, respectively. Instead, the extremal pathways correspond to snaking through the CRN, alternately using both the template and destructive catalyst to first create a sequence, and then convert it into another sequence. We show these pathways for $L = 4$ in appendix G. These SAWs would not only be absurd as

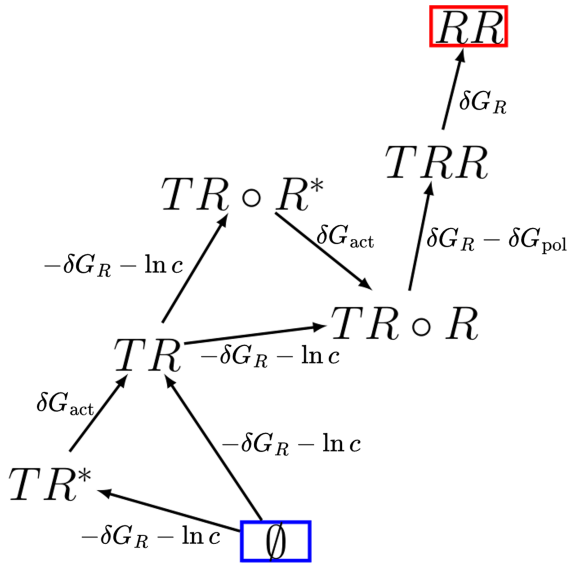


FIG. 8: Modifying the template-based reactions of the model in section V A to include kinetic proofreading. We show here the template-based reactions leading directly to the dimer RR, which should replace the equivalent template-based reactions in figure 6. As before, arrows represent reversible reactions with free-energy change in the direction of the arrow indicated. On the template, from state TR, either a non-activated (R^*) or activated (R) monomer may bind to the template. If that monomer has bound, but has not yet been polymerised into the growing polymer, it is represented by $TR \circ R^*$ or $TR \circ R$. When bound to the template, non-activated monomers may be activated, as shown by the transitions in which R^* is converted to R. When there is an activated monomer at the end of the growing polymer ($TR \circ R$), that monomer may be polymerised into the growing polymer to reach a polymerised state (TRR). After a full length polymer has grown on the template, it may detach to a product (RR for a dimer template).

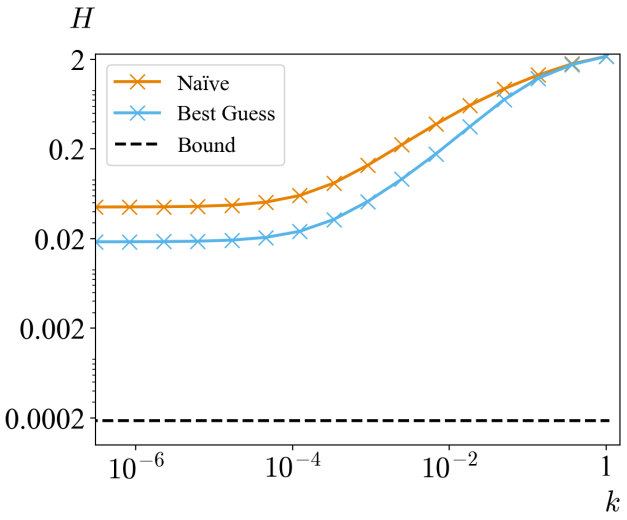


FIG. 9: Entropy H for two attempts to optimise the entropy, plotted alongside the lower bound $H_{\rm min}$ implied by eqs. 20 and 21, plotted as a function of he parameter k that sets the overall scale of slow reactions relative to fast ones. In the "naive" approach, rates favour pathways to the correct product on the template and the incorrect ones on the destroyer. The "best guess" favours the snaking pathways described in text. The data is obtained for $\delta G_f = 1$, $\delta G_{\rm act} = 1$, $\delta G_{\rm pol} = 0$, $\delta G_R = 2$, $\delta G_W = -2$ and $c = 1$.

the dominant pathways in a real system, they actually define a bound that is unachievable, even in principle. Since the pathways exhibiting δG_U and δG_L pass through other products as intermediates, scaling these pathways to be fast would necessarily result in sub-pathways to other products being fast.

Nonetheless, the existence of these snaking pathways can provide some advantage, at least in principle. In figure 9, we show two attempts to find parameters that minimize entropy for a system with $L = 4$. In the first "naïve" scheme, plotted in blue, reactions contributing to assembly of $RRRR$ via the template or assembly of any other sequence via the destructive catalyst are assigned rates of 1 and all other rates are taken as $\sim k$. We show this scheme in appendix G. In the second "best guess" scheme, plotted in red, we make use of these snaking pathways. The reactions contributing to assembly of $RRRR$ via the template are still assigned rates of 1. However, for all other products, the longest snaking pathway which does not intersect with the pathway leading to $RRRR$ has rates assigned 1. Again we show this scheme in more detail in appendix G.

As $k \to 0$, the snaking pathways outperform the non-snaking pathways significantly, resulting in approximately half the entropy of the product distribution (figure 9). Notably, however, neither the best guess nor the naïve system converges on the bound as $k \to 0$; the best guess approaches $H = 1.8 \times 10^{-2}$ as compared to the bound at $H_{\rm min} = 1.9 \times 10^{-4}$, two orders of magnitude lower.

Although the the best guess system can outperform the naïve system in principle, when absolute rates can be chosen freely, rates may in practice be mechanistically constrained. As a result, this outperformance may not be achievable in a specific model. Indeed, at moderate k, the entropies of the naive system and the best guess converge.

VI. CONCLUSION

Considering systems in which catalysts selectively produce products drawn from a large set of possibilities, we have derived general bounds on the steady-state distribu-

tion of products. The maximal difference in free energies between formation paths for the products sets an upper bound on the specificity (or probability with which a single sequence can be selected from the product distribution) and a lower bound on the entropy of that product distribution. These bounds are remarkably simple, applying to arbitrarily complex networks regardless of details such as kinetic proofreading.

Several features of these bounds are, a least to us, unexpected. They modify our understanding that producing sequence-specific products has a "cost" related to the information stored [2, 31, 32], and help to resolve the paradox that thermodynamics imposes limits on templating despite the fact that catalysts can be arbitrarily specific for their substrate with no minimal cost.

Most importantly, the bounds are pseudo-equilibrium in nature. In the optimal system, kinetic selectivity ensures that each product is coupled to only a single pathway, with the yield determined by the equilibrium properties of that pathway. As a result, although maintaining an ensemble with exactly one product drawn from M possibilities requires pathways differentiated by ${}^{\Delta G}/_{\ln M} > 1$, this ΔG is not a cost in the sense of entropy production. A low entropy product distribution will be far from equilibrium, with a high non-equilibrium free energy, and will consequently require more chemical work to create it in the first place than a uniform ensemble [2, 29]. But once in the dynamic steady state, a highly-specific ensemble (which is capable of dynamically responding to changing catalyst concentrations) can, in theory, be maintained with negligible "housekeeping" entropy production [45].

Such behaviour is profoundly different from apparently similar information-processing networks, such as a push-pull phosphorylation network [46]. In that case, accurate information transmission is only possible if the system undergoes cyclic phosphorylation by one pathway, and dephosphorylation by another, resulting in a high housekeeping entropy production. Fundamentally, the difference arises because the system considered here relies on catalytic rates distinguishing between different pathways to different products, which is not constrained by entropy production, whereas the push-pull network of [46] relies on catalysts selecting a forwards pathway over its reverse, which is so constrained [33].

Indeed, the thermodynamic limits on the product distribution that do exist do not reflect the actual selectivity of the underlying catalytic processes, which is the typical focus of work on templated processes [6–22, 25–30]. The ability of a template to kinetically select certain products determines how close to the bound a system can come, but it does not determine the bound itself. For example, although a single product cannot dominate the steady state for ${}^{\Delta G}/_{\ln M} < 1$, this does not mean that catalysts cannot be arbitrarily specific; simply that the steady state achieved via the set of available catalysts cannot be dominated by one product. Indeed, remarkably, for $M \to \infty$ the lower bound on product entropy is achieved when a vanishingly small fraction $m_{\min}/M$ of the possible products dominate the ensemble, even for $\Delta G \ll \ln M$. Such a distribution could arise from a system with $m_{\min}$ templates each catalysing a single product with high accuracy, and a non-specific destructive catalyst.

The above results suggest that, in some sense, catalytic templating ensembles do not have an inherent minimum maintenance "cost" related to the accuracy of information propagation from a single template to its product. From another perspective, however, our bounds show that maintaining a product distribution that is highly-specific for a single product is hard relative to previous claims [2, 31, 32]. The minimal entropy distribution is only dominated by a single product for $\Delta G \gtrsim \ln M + \ln\ln M$, a condition that converges to the more familiar ${}^{\Delta G}/_{\ln M} > 1$ achingly slowly with M. Moreover, the minimal entropy of a product distribution with finite M is larger than that for $M \to \infty$ for all values of ${}^{\Delta G}/_{\ln M}$. For any finite M, perfect accuracy is only possible in the limit $(\Delta G - \ln M) \to \infty$, although good accuracy is possible for $\Delta G \gtrsim \ln M + \ln\ln M$.

This work open up several lines of inquiry. Firstly: why do cells not operate in the manner suggested by the optimal behaviour? RNA molecules are not destroyed by the reverse of their creation pathway. As we have noted, the bounds may be formally unachievable even in a system for which rates can be arbitrarily scaled, but even in this context it is theoretically possible to produce sharply-peaked ensembles at low cost via mechanisms that avoid cycling of products via distinct pathways. We expect that practical constraints that would limit the speed and specificity of catalysts operating in this reversible regime must be a factor. It may also be true that, in practical contexts, the "excess" entropy production due to responding to changing conditions [45] dominates of the housekeeping entropy production for the steady state. Indeed, exploring how the behaviour of realistic models that are constrained by the actual chemistry and biological context would be highly informative. We note, however, that synthetic information-processing systems might be engineered to operate in this low-cost fashion - particularly if the products are relatively simple compared to proteins and RNA. Such a system would be reminiscent of the behaviour of dynamic combinatorial chemistry ensembles [47], but with the system designed so that the presence/absence of specific catalysts changes the free energy of the dominant pathway to an assembled product.

Secondly, we have made a number of simplifying assumptions in this work. Most significantly, we have assumed that all products are equally thermodynamically stable, and produced via equivalent pathways. In practice, although the presence of catalysts cannot change the thermodynamic stability of the products, some products will be more stable than others due to intra-product interactions, and network topologies will necessarily vary for polymers of different length. As a result, δG_U and δG_L will be product-specific in general. In this case,

much of our analysis would still apply directly, with individual products pushed either to their maximum or minimum concentration to minimize entropy, and the optimal distribution being achieved when each product is overwhelmingly coupled to a single pathway. However, it will likely be easier to create distributions that are sharply peaked about the most stable products, and harder in other cases. Given that the aim of processes like transcription and translation is to produce functional, rather than stable, products, it is unclear whether this asymmetry would actually be beneficial. It would also be interesting to examine behaviour when either monomers or catalysts are not chemostatted, making the system nonlinear.

Finally, we have not considered the dynamics of these networks in detail. We have shown that at finite times, the bounds can be violated, but we have not attempted to explore for how long this is possible. Indeed, one might ask whether cells are effectively in this non-steady-state limit since some theoretically possible polymers (such as chimeric RNA/DNA sequences) are essentially never created by the cell. The question of how these networks respond to dynamical changes in the concentration of the catalysts, and the associated thermodynamic costs, is also open.

Author Contributions

All authors conceived of the project. B.Q. produced the analysis and wrote the initial draft. All authors interpreted results and reviewed and edited this paper.

DATA AVAILABILITY STATEMENT

The Mathematica code that support the findings of this study openly available on Zenodo at `https://doi.org/10.5281/zenodo.10909084`.

ACKNOWLEDGMENTS

We thank Pieter Rein ten Wolde for his constructive comments on the manuscript. This work is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant agreement No. 851910). T.E.O. is supported by a Royal Society University Fellowship.

Appendix A: Minimizing entropy maximizes channel capacity

We can consider the deterministic CRNs in this paper to be information channels. Let us assume the underlying chemistry, which determines the rate constants appearing in the un-linearized model, and the concentration of monomers, are fixed. The effective rates of the linearized network would then vary with the concentrations of catalysts only. As a specific (simple) example of how the entropy bound defines the channel capacity, let us assume that all variability is due to M sequence-specific template catalysts, one for each product, and that of these templates exactly one is present at any given time at a fixed concentration.

We can define the input to the information channel as the template that is present at high concentration; the output would then be a product sampled from the steady state product distribution for that input state. If the templates are symmetric, acting equivalently relative to their ideal sequence, then each output distribution would merely be a permutation of the set of product probabilities.

A system of this kind would define a symmetric channel. The channel capacity of such a symmetric channel [34] is given by $C = \ln M - H([p_i])$ in our notation, where M is the number of products/sequence-specific templates, and $H([p_i])$ the Shannon entropy of the output distribution for any single template. This entropy will be the same for any input state in our symmetrized description. Hence, minimizing the entropy maximizes the channel capacity.

Appendix B: Proof of the steady state concentration written as sum over spanning trees

We note that a similar result can be seen in *e.g.* [39]. However, we prove this result here with particular reference to the effect of a null species.

Consider a linear connected CRN under mass action kinetics. Let X_i be the chemical species with concentration c_i. Assume the CRN contains some reactions of the form $\emptyset \rightleftharpoons X$. Without loss of generality, we assume that there only exists at most one reaction of the form $X_i \to X_j$ (for the cases where there are multiple such reactions, replace their reaction rate with the sum over all reaction rates of reactions of that form). We may cast the steady state equation for the vector of steady state concentrations of the chemical species, $\mathbf{c}$, as the linear equation:

$$A\mathbf{c} + \mathbf{b} = \mathbf{0}, \tag{B1}$$

where $\mathbf{b}$ is a vector such that entry b_i is the reaction rate of the reaction $\emptyset \to X_i$, and A is a matrix with off diagonal entries A_{ij} equal to the reaction rate of the reaction $X_j \to X_i$ and diagonal elements A_{ii} equal to minus the sum of reaction rates of reactions $X_i \to X_j$ over all chemical species X_j $j \neq i$ as well as $X_i \to \emptyset$. Note that the sum of column i of matrix A is equal to minus the reaction rate of reaction $X_i \to \emptyset$. Thus, we

may create the new matrix:

$$K = \left(\begin{array}{c|c} A & \mathbf{b} \\ \hline \mathbf{d}^T & -\sum_i b_i \end{array}\right), \tag{B2}$$

where $\mathbf{d}$ is a vector such that entry d_i is equal to the reaction rate of reaction $X_i \to \emptyset$. The columns of the matrix, K, now sum to zero and we can use the all minors matrix tree theorem [48]. Represent K as a graph with nodes corresponding to chemical species X_i and an additional node corresponding to $\emptyset$, and edges e corresponding to reactions between species with weights equal to their reaction rate constants $k(e)$. Then, the determinant (up to a sign) of the sub matrix formed by deleting row and column i from matrix K is given by the sum over the set of spanning trees rooted at node i, $\mathcal{T}(X_i)$,

$$\det(K/i) = \sum_{T\in\mathcal{T}(X_i)} \prod_{e\in T} k(e), \tag{B3}$$

where (K/i) represents matrix K with row and column i deleted. In particular,

$$\det(A) = \sum_{T\in\mathcal{T}(\emptyset)} \prod_{e\in T} k(e). \tag{B4}$$

Thus, by Cramers rule [49], and the sign of the determinant under swapping of rows, eq. 2 gives the solution to eq. B1.

Appendix C: Proof of boundedness of steady-state concentrations

We note that similar proofs exist in the literature [39–42], but we have included the proof here for completeness and to match our specific conventions. Further, in the cases considered in this paper, the fact that all products are connected to the null state means that our result bounds the absolute, rather than a relative, concentrations.

Consider a linear connected CRN with chemical species X_i and some reactions of the form $\emptyset \rightleftharpoons X_i$. The concentration of species X_i may be written as in eq 2. The numerator and denominator in this fraction are sums over spanning trees rooted at a given node. A sum over spanning trees rooted at node X may be factored into a sum over self avoiding walks (SAWs) from some arbitrary other node to node X. Concretely, letting Y be the other, arbitrary node, and $S(Y \to X)$ be the set of SAWs from Y to X,

$$\sum_{T\in\mathcal{T}(X)} \prod_{e\in T} k(e) = \sum_{S\in S(Y\to X)} A(S) \prod_{e\in S} k(e), \tag{C1}$$

where $A(S)$ is a factor that, crucially, is the same for the equivalent (reversed) SAW in $S(X \to Y)$, in which all edges are reversed compared to $S(Y \to X)$. That is to say, if we now wish to find the sum over spanning trees rooted at Y, we may choose X as the arbitrary other state and find:

$$\sum_{T\in\mathcal{T}(Y)} \prod_{e\in T} k(e) = \sum_{S\in S(Y\to X)} A(S) \prod_{e\in S} k(\bar{e}), \tag{C2}$$

where $\bar{e}$ is the reverse of edge e. For the linear CRNs, utilising eq 3, we may thus write:

$$\begin{aligned} c_X &= \frac{\sum_{S\in S(\emptyset\to X)} A(S) \prod_{e\in S} k(e)}{\sum_{S\in S(\emptyset\to X)} A(S) \prod_{e\in S} k(\bar{e})} \\ &= \frac{\sum_{S\in S(\emptyset\to X)} A(S) \left[\prod_{e\in S} k(\bar{e})\right] e^{-\delta G_S}}{\sum_{S\in S(\emptyset\to X)} A(S) \prod_{e\in S} k(\bar{e})}. \end{aligned} \tag{C3}$$

Hence,

$$c_X \in \left[e^{-\max_{S\in S(\emptyset\to X)}(\delta G_S)}, e^{-\min_{S\in S(\emptyset\to X)}(-\delta G_S)}\right], \tag{C4}$$

as required.

Appendix D: Proof of the boundedness of steady-state distribution entropy

We have a set of M concentrations $\{c_1, \dots c_M\}$. Denote the total concentration $c_T = \sum_{i=1}^{M} c_i$. From these concentrations we can define a distribution $\{p_1, \dots p_M\}$ where $p_i = \frac{c_i}{c_T}$. The Shannon entropy of this distribution is

$$H([p_i]) = -\sum_{i=1}^{M} p_i \ln p_i = -\frac{1}{c_T}\sum_{i=1}^{M} c_i \ln c_i + \ln c_T. \tag{D1}$$

Suppose that each concentration is bounded by the same upper and lower bounds, $c_i \in [c_L, c_U]$. We now propose that the distribution of concentrations that minimises the entropy, $H([c_i])$, is that with $m_{\min}$ of the species at concentration c_U and $M - m_{\min}$ at concentration c_L, where $m_{\min}$ is either

$$\left\lceil \frac{\frac{c_L}{c_U}\left[-\ln\left(\frac{c_L}{c_U}\right) - \left(1-\frac{c_L}{c_U}\right)\right]}{\left(1-\frac{c_L}{c_U}\right)^2} M \right\rceil \text{ or}$$
$$\left\lceil \frac{\frac{c_L}{c_U}\left[-\ln\left(\frac{c_L}{c_U}\right) - \left(1-\frac{c_L}{c_U}\right)\right]}{\left(1-\frac{c_L}{c_U}\right)^2} M \right\rceil - 1. \tag{D2}$$

To prove this claim, let us calculate the derivative of $H = H([p_i])$ with respect to a concentration c_α, holding

all other concentrations fixed and remembering that c_T is linear in c_α,

$$\frac{\partial H}{\partial c_\alpha} = \frac{1}{c_T}\left(-\ln\left(\frac{c_\alpha}{c_T}\right) - H\right). \tag{D3}$$

Additionally, we require the second derivative,

$$\frac{\partial^2 H}{\partial c_\alpha^2} = -\frac{1}{c_T^2}\left(-\ln\left(\frac{c_\alpha}{c_T}\right) - H\right) - \frac{1}{c_T}\frac{\partial H}{\partial c_\alpha} - \frac{1}{c_T}\frac{1}{c_\alpha}\left(1 - \frac{c_\alpha}{c_T}\right) \tag{D4}$$

Evaluating the second derivative at $-\ln\left(\frac{c_\alpha}{c_T}\right) = H$, the first two terms are zero and the third is necessarily negative for non-zero c_L. Thus

$$\left.\frac{\partial^2 H}{\partial c_\alpha^2}\right|_{-\ln\left(\frac{c_\alpha}{c_T}\right)=H} < 0. \tag{D5}$$

And so, for any distribution, we can decrease the H by increasing the concentrations of any species i for which $-\ln\left(\frac{c_i}{c_T}\right) < H$ and decreasing the concentrations for any species whose concentration has $-\ln\left(\frac{c_i}{c_T}\right) > H$. For species for which $-\ln\left(\frac{c_i}{c_T}\right) = H$, changing the concentration in either direction will decrease H. Consequently, minimising the entropy of the distribution necessarily requires all species to be at one bound or the other: we need m species at concentration c_U and $M-m$ at concentration c_L. Hence, we have transformed the problem into a one dimensional one of minimising H as a function of m. We can write this entropy as

$$H(m) = -\frac{(M-m)\frac{c_L}{c_U}\ln\left(\frac{c_L}{c_U}\right)}{(M-m)\frac{c_L}{c_U}+m} + \ln\left((M-m)\frac{c_L}{c_U}+m\right). \tag{D6}$$

Taking the derivative with respect to m and setting it to zero tells us that

$$m_{\min} = \frac{\frac{c_L}{c_U}\left[-\ln\left(\frac{c_L}{c_U}\right) - \left(1-\frac{c_L}{c_U}\right)\right]}{\left(1-\frac{c_L}{c_U}\right)^2} M \tag{D7}$$

gives a turning point for $H(m)$, which is clearly a minimum since $H(0) = H(M) = \ln M$ is maximal entropy. Since $H(m)$ has only a single turning point as a function of m, the integer that minimises $H(m)$ will be either the floor or ceiling of the above expression, and $H_{\min}$ is given by eqs. 12 and 13.

Appendix E: Chemical reaction network for example 1

We present here the full chemical reaction network used for section V A.

$$\begin{aligned}
T + X_1 &\underset{k^T_{X_1} e^{-\delta G_{X_1}}}{\overset{k^T_{X_1}}{\rightleftharpoons}} TX_1 \\
TX_1\cdots X_{n-1} + X_n &\underset{k^T_{X_1\cdots X_n} e^{-\delta G_{X_n} - \delta G_{\rm pol}}}{\overset{k^T_{X_1\cdots X_n} e^{-\delta G_{X_{n-1}}}}{\rightleftharpoons}} TX_1\cdots X_n \\
TX_1\cdots X_L &\underset{k^{T,u}_{X_1\cdots X_L}}{\overset{k^{T,u}_{X_1\cdots X_L} e^{-\delta G_{X_L}}}{\rightleftharpoons}} T + X_1\cdots X_L \\
D + X_1 &\underset{k^D_{X_1} e^{\delta G_f - \delta G_{X_1}}}{\overset{k^D_{X_1}}{\rightleftharpoons}} DX_1 \\
DX_1\cdots X_{n-1} + X_n &\underset{k^D_{X_1\cdots X_n} e^{\delta G_f - \delta G_{X_n} - \delta G_{\rm pol}}}{\overset{k^D_{X_1\cdots X_n} e^{-\delta G_{X_{n-1}}}}{\rightleftharpoons}} DX_1\cdots X_n \\
DX_1\cdots X_L &\underset{k^{D,u}_{X_1\cdots X_L}}{\overset{k^{D,u}_{X_1\cdots X_L} e^{-\delta G_{X_L}}}{\rightleftharpoons}} D + X_1\cdots X_L,
\end{aligned} \tag{E1}$$

for $n \le L$, $X_i \in \{R, W\}$. Dynamics is assumed to follow mass action kinetics, with rate constants given above and below the harpoons. T represents the template, D the destructive catalyst and R, W the "right" and "wrong"

monomers. The products are $X_1 \cdots X_L$, representing the polymers of length L. Here, there are $M = 2^L$ different products. $TX_1 \cdots X_n$ and $DX_1 \cdots X_n$ represent partial polymers, $X_1 \cdots X_n$, bound to the template or destructive catalyst. Monomer X binds to the template or destructive catalyst with standard free-energy $-\delta G_X$ and the standard free-energy of polymerisation is $-\delta G_{\rm pol}$ in the absence of fuel. The destructive catalyst has an additional free energy δG_f per length driving the disassembly of polymers.

Let us consider a set of rate constants that allow us to saturate (in a certain limit) the bound for specificity maximisation. We may for example, set the rates of $\emptyset \to TR$, $TR \to TRR$, ..., $TR^{L-1} \to TR^L$, $TR^L \to R^L$ equal to a non-zero constant $k_J = 1$ (here, R^L corresponds to L copies of R). We also set all the rates of $\emptyset \to DX_1$, $DX_1 \to DX_1X_2$, ..., $DX_1...X_{L-1} \to DX1...X_L$, $DX_1...X_L \to X_1..X_L$, where $X_i = R$ or W but excluding $X_1 \cdots X_L$ all being R, equal to 1. Conversely, we set the rates of $\emptyset \to DR$, $DR \to DRR$, ..., $DR^{L-1} \to DR^L$, $DR^L \to R^L$ and $\emptyset \to TX_1$, $TX_1 \to TX_1X_2$, ..., $TX_1...X_{L-1} \to TX_1...X_L$, $TX_1...X_L \to X_1..X_L$, where $X_i = R$ or W but excluding $X_1 \cdots X_L$ all being R, equal to $k_I = k$. The reverse reactions of those listed above have a rate determined by the free-energy change of reaction. For $k \to 0$, this set of reaction rates saturates the bound.

Appendix F: Chemical reaction network for example 2

Similarly, we present the full CRN used for section V B.

$$
\begin{aligned}
T + X_1^* &\underset{k^{T,I}_{X_1} e^{-\delta G_{X_1}}}{\overset{k^{T,I}_{X_1}}{\rightleftharpoons}} TX_1^* \\
TX_1^* &\underset{k^{T,\mathrm{act}}_{X_1}}{\overset{k^{T,\mathrm{act}}_{X_1} e^{\delta G_{\mathrm{act}}}}{\rightleftharpoons}} TX_1 \\
T + X_1 &\underset{k^{T}_{X_1} e^{-\delta G_{X_1}}}{\overset{k^{T}_{X_1}}{\rightleftharpoons}} TX_1 \\
TX_1 \cdots X_{n-1} + X_n^* &\underset{k^{T,I}_{X_1 \cdots X_n} e^{-\delta G_{X_n}}}{\overset{k^{T,I}_{X_1 \cdots X_n}}{\rightleftharpoons}} TX_1 \cdots X_{n-1} \circ X_n^* \\
TX_1 \cdots X_{n-1} \circ X_n^* &\underset{k^{T,\mathrm{act}}_{X_1 \cdots X_n}}{\overset{k^{T,\mathrm{act}}_{X_1 \cdots X_n} e^{\delta G_{\mathrm{act}}}}{\rightleftharpoons}} TX_1 \cdots X_{n-1} \circ X_n \\
TX_1 \cdots X_{n-1} + X_n &\underset{k^{T}_{X_1 \cdots X_n} e^{-\delta G_{X_n}}}{\overset{k^{T}_{X_1 \cdots X_n}}{\rightleftharpoons}} TX_1 \cdots X_{n-1} \circ X_n \\
TX_1 \cdots X_{n-1} \circ X_n &\underset{k^{T,\mathrm{pol}}_{X_1 \cdots X_n} e^{-\delta G_{\mathrm{pol}}}}{\overset{k^{T,\mathrm{pol}}_{X_1 \cdots X_n} e^{-\delta G_{X_{n-1}}}}{\rightleftharpoons}} TX_1 \cdots X_n \\
TX_1 \cdots X_L &\underset{k^{T,u}_{X_1 \cdots X_L}}{\overset{k^{T,u}_{X_1 \cdots X_L} e^{-\delta G_{X_L}}}{\rightleftharpoons}} T + X_1 \cdots X_L \\
D + X_1 &\underset{k^{D}_{X_1} e^{\delta G_f - \delta G_{X_1}}}{\overset{k^{D}_{X_1}}{\rightleftharpoons}} DX_1 \\
DX_1 \cdots X_{n-1} + X_n &\underset{k^{D}_{X_1 \cdots X_n} e^{\delta G_f - \delta G_{X_n} - \delta G_{\mathrm{pol}}}}{\overset{k^{D}_{X_1 \cdots X_n} e^{-\delta G_{X_{n-1}}}}{\rightleftharpoons}} DX_1 \cdots X_n \\
DX_1 \cdots X_L &\underset{k^{D,u}_{X_1 \cdots X_L}}{\overset{k^{D,u}_{X_1 \cdots X_L} e^{-\delta G_{X_L}}}{\rightleftharpoons}} D + X_1 \cdots X_L,
\end{aligned}
\tag{F1}
$$

where $n \leq L, X_i \in \{R, W\}$, $X_i^* \in \{R^*, W^*\}$. As in the previous CRN, T represents the template, D the

destructive catalyst, R the "right" monomers and W the "wrong" monomers. R^* and W^* are non-activated monomers. Dynamics is assumed to follow mass action kinetics, with rate constants given above and below the harpoons. The species, $TX_1 \cdots X_{n-1} \circ X_n$ ($TX_1 \cdots X_{n-1} \circ X_n^*$) represent a complex of polymer of length $n-1$ bound to the template as well as a (non-)activated monomer X_n (X_n^*) bound, but not yet polymerised into a single polymer of length n. $X_1 \cdots X_L$ are the products. Both non-activated (X^*) and activated monomers (X) bind to the template or destructive catalyst with standard free energy $-\delta G_X$. If an non-activated monomer is bound to the template, it may be activated, with a free-energy change δG_{act}. If an activated monomer is bound to the template, it may be polymerised into the growing copolymer; the standard free-energy change of polymerisation is $-\delta G_{\text{pol}}$. The destructive catalyst has an additional free-energy δG_f per length driving the disassembly of polymers.

Appendix G: Pathways for example 2

In this appendix, for the kinetic proofreading CRN (section V B) we depict the pathways of most positive/negative free-energy change (figure 10) as well as the fast reactions used for the two attempts to minimize the entropy, the naïve attempt (figure 11) and our best guess (figure 12). For simplicity, we have coarse-grained the intermediate states for the reactions involving templates, and have presented these processes using two lines for the pathways through the kinetic proofreading motifs. The solid line represents the inactivated monomer pathway ($TX_1 \cdots X_{n-1} \to TX_1 \cdots X_{n-1} \circ X_n^* \to TX_1 \cdots X_{n-1} \circ X_n \to TX_1 \cdots X_n$) and the dashed line represents the active monomer pathway ($TX_1 \cdots X_{n-1} \to TX_1 \cdots X_{n-1} \circ X_n \to TX_1 \cdots X_n$). Extremal pathways (figure 10) and fast reactions with rates ~ 1 (figure 11 and figure 12) are highlighted in red.

[1] F. Crick, Central dogma of molecular biology, Nature **227**, 561 (1970).

[2] T. E. Ouldridge and P. R. ten Wolde, Fundamental costs in the production and destruction of persistent polymer copies, Physical Review Letters **118**, 158103 (2017).

[3] J. D. Watson, *Molecular biology of the gene*, seventh edition. ed. (Benjamin-Cummings Publishing Company, Boston, 2014).

[4] J. J. Hopfield, Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity, Proceedings of the National Academy of Sciences of the United States of America **71**, 4135 (1974).

[5] J. Ninio, Kinetic amplification of enzyme discrimination, Biochimie **57**, 587 (1975).

[6] C. H. Bennett, Dissipation-error tradeoff in proofreading, BioSystems **11**, 85 (1979).

[7] A. Murugan, D. A. Huse, and S. Leibler, Discriminatory proofreading regimes in nonequilibrium systems, Physical Review X **4**, 021016 (2014).

[8] A. Murugan, D. A. Huse, and S. Leibler, Speed, dissipation, and error in kinetic proofreading, Proceedings of the National Academy of Sciences of the United States of America **109**, 12034 (2012).

[9] Q. Yu, A. B. Kolomeisky, and O. A. Igoshin, The energy cost and optimal design of networks for biological discrimination, Journal of the Royal Society Interface **19**, 20210883 (2022).

[10] M. Sahoo and S. Klumpp, Backtracking dynamics of rna polymerase: pausing and error correction, Journal of Physics: Condensed Matter **25**, 374104 (2013).

[11] M. Sahoo, N. Arsha, P. R. Baral, and S. Klumpp, Accuracy and speed of elongation in a minimal model of dna replication, Physical Review E **104**, 034417 (2021).

[12] M. Ehrenberg and C. Blomberg, Thermodynamic constraints on kinetic proofreading in biosynthetic pathways, Biophysical Journal **31**, 333 (1980).

[13] Y.-S. Song, Y.-G. Shu, X. Zhou, Z.-C. Ou-Yang, and M. Li, Proofreading of dna polymerase: a new kinetic model with higher-order terminal effects, Journal of Physics: Condensed Matter **29**, 025101 (2016).

[14] Y. Song and C. Hyeon, Thermodynamic cost, speed, fluctuations, and error reduction of biological copy machines, The Journal of Physical Chemistry Letters **11**, 3136 (2020).

[15] Q.-S. Li, P.-D. Zheng, Y.-G. Shu, Z.-C. Ou-Yang, and M. Li, Template-specific fidelity of dna replication with high-order neighbor effects: A first-passage approach, Physical Review E **100**, 012131 (2019).

[16] F. Wong, A. Amir, and J. Gunawardena, Energy-speed-accuracy relation in complex networks for biological discrimination, Physical Review E **98**, 012420 (2018).

[17] W. D. Piñeros and T. Tlusty, Kinetic proofreading and the limits of thermodynamic uncertainty, Physical Review E **101**, 022415 (2020).

[18] M. Nguyen and S. Vaikuntanathan, Design principles for nonequilibrium self-assembly, Proceedings of the National Academy of Sciences of the United States of America **113**, 14231 (2016).

[19] V. Galstyan and R. Phillips, Allostery and kinetic proofreading, The Journal of Physical Chemistry B **123**, 10990 (2019).

[20] R. Rao and L. Peliti, Thermodynamics of accuracy in kinetic proofreading: dissipation and efficiency trade-offs, Journal of Statistical Mechanics: Theory and Experiment **2015**, P06001 (2015).

[21] K. Banerjee, A. B. Kolomeisky, and O. A. Igoshin, Elucidating interplay of speed and accuracy in biological error correction, Proceedings of the National Academy of Sciences of the United States of America **114**, 5183 (2017).

[22] D. Chiuchiù, Y. Tu, and S. Pigolotti, Error-speed correlations in biopolymer synthesis, Physical Review Letters **123**, 038101 (2019).

[23] P. Sartori and S. Pigolotti, Kinetic versus energetic discrimination in biological copying, Physical Review Letters **110**, 188101 (2013).

[24] P. Gaspard and D. Andrieux, Kinetics and thermodynamics of first-order markov chain copolymerization, The Journal of Chemical Physics **141**, 044908 (2014).

[25] J. M. Poulton, P. R. Ten Wolde, and T. E. Ouldridge, Nonequilibrium correlations in minimal dynamical models of polymer copying, Proceedings of the National Academy of Sciences **116**, 1946 (2019).

[26] S. Pigolotti and P. Sartori, Protocols for copying and proofreading in template-assisted polymerization, Journal of Statistical Physics **162**, 1167 (2016).

[27] P. Sartori and S. Pigolotti, Thermodynamics of error correction, Physical Review X **5**, 041039 (2015).

[28] B. Qureshi, J. Juritz, J. M. Poulton, A. Beersing-Vasquez, and T. E. Ouldridge, A universal method for analyzing copolymer growth, The Journal of Chemical Physics **158**, 104906 (2023).

[29] J. M. Poulton and T. E. Ouldridge, Edge-effects dominate copying thermodynamics for finite-length molecular oligomers, New Journal of Physics **23**, 063061 (2021).

[30] J. Juritz, J. M. Poulton, and T. E. Ouldridge, Minimal mechanism for cyclic templating of length-controlled copolymers under isothermal conditions, The Journal of Chemical Physics **156**, 074103 (2022).

[31] C. H. Bennett, The thermodynamics of computation—a review, International Journal of Theoretical Physics **21**, 905 (1982).

[32] A. Genthon, C. D. Modes, F. Jülicher, and S. W. Grill, Non equilibrium transitions in a polymer replication ensemble, arXiv preprint arXiv:2403.05665 (2024).

[33] T. E. Ouldridge, The importance of thermodynamics for molecular systems, and the importance of molecular systems for thermodynamics, Natural Computing **17**, 3 (2018).

[34] T. M. Cover and J. A. Thomas, *Elements of information theory*, second edition. ed. (Wiley-Interscience, Hoboken, N.J, 2006).

[35] C. E. Shannon, A mathematical theory of communication, The Bell system technical journal **27**, 379 (1948).

[36] M. Polettini and M. Esposito, Irreversible thermodynamics of open chemical networks. i. emergent cycles and broken conservation laws, The Journal of Chemical Physics **141**, 024117 (2014).

[37] R. Rao and M. Esposito, Nonequilibrium thermodynamics of chemical reaction networks: Wisdom from stochastic thermodynamics, Phys. Rev. X **6**, 041064 (2016).

[38] D. F. Anderson, G. Craciun, and T. G. Kurtz, Product-form stationary distributions for deficiency zero chemical reaction networks, Bulletin of Mathematical Biology **72**, 1947 (2010).

[39] K.-M. Nam, R. Martinez-Corral, and J. Gunawardena, The linear framework: using graph theory to reveal the algebra and thermodynamics of biomolecular systems, Interface Focus **12**, 20220013 (2022).

[40] U. Çetiner and J. Gunawardena, Reformulating nonequilibrium steady states and generalized hopfield discrimination, Physical Review E **106**, 064128 (2022).

[41] C. Maes and K. Netočnỳ, Heat bounds and the blowtorch theorem, in *Annales Henri Poincaré*, Vol. 14 (Springer, 2013) pp. 1193–1202.

[42] M. Sáez, E. Feliu, and C. Wiuf, Linear elimination in chemical reaction networks, in *Recent Advances in Differential Equations and Applications* (Springer International Publishing, Cham, 2019) pp. 177–193.

[43] P. C. Nelson, *Biological physics: energy, information, life* (Chiliagon Science, Philadelphia PA USA, 2020).

[44] J. D. Mallory, O. A. Igoshin, and A. B. Kolomeisky, Do we understand the mechanisms used by biological systems to correct their errors?, The Journal of Physical Chemistry B **124**, 9289 (2020).

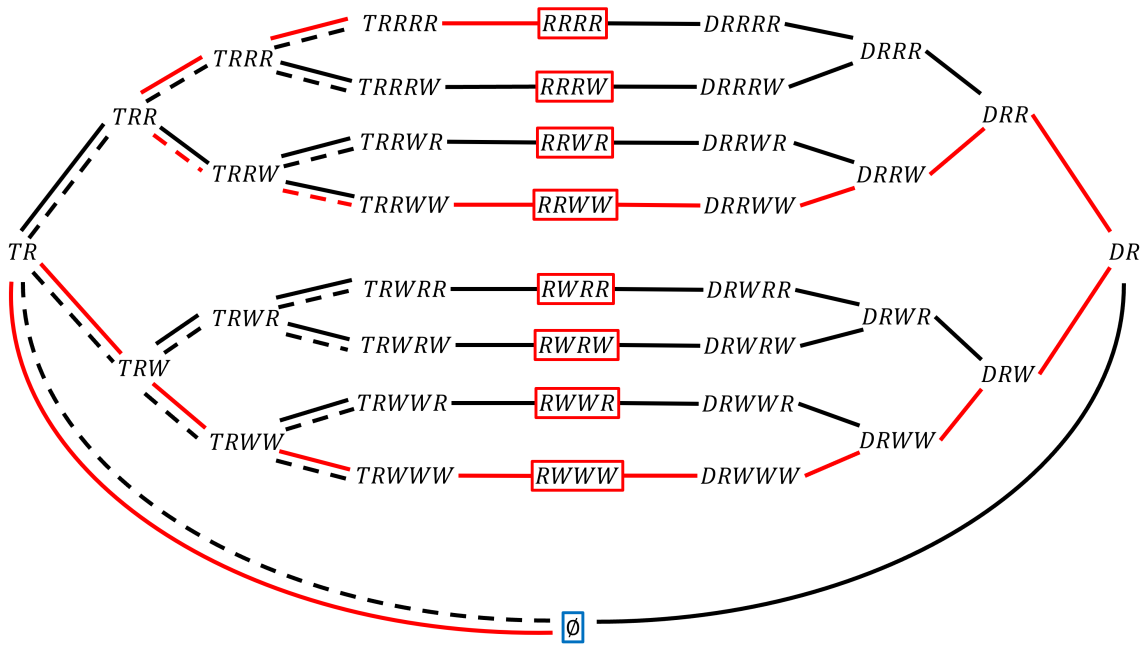
[45] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, Reports on progress in physics **75**, 126001 (2012).

[46] T. E. Ouldridge, C. C. Govern, and P. R. ten Wolde, Thermodynamics of computational copying in biochemical systems, Phys. Rev. X **7**, 021004 (2017).

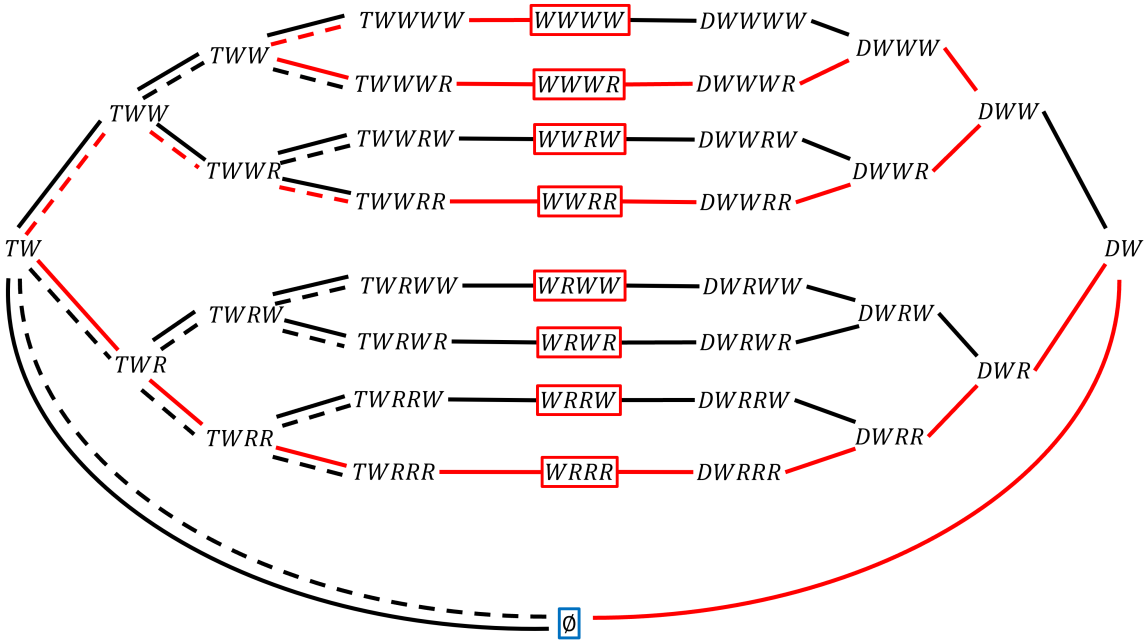
[47] S. Ladame, Dynamic combinatorial chemistry: on the road to fulfilling the promise, Org. Biomol. Chem. **6**, 219 (2008).

[48] S. Chaiken, A combinatorial proof of the all minors matrix tree theorem, SIAM Journal on Algebraic Discrete Methods **3**, 319 (1982).

[49] S. M. Robinson, A short proof of cramer's rule, Mathematics Magazine **43**, 94 (1970).



(a)



(b)

FIG. 10: (a) The pathway to $RRRR$ with free energy change δG_U given by eq. 20 is shown in red. (b) The pathway to $WWWW$ with free energy change δG_L given by eq. 21 is shown in red. For each of these diagrams, only the relevant half the reaction network is shown for simplicity.

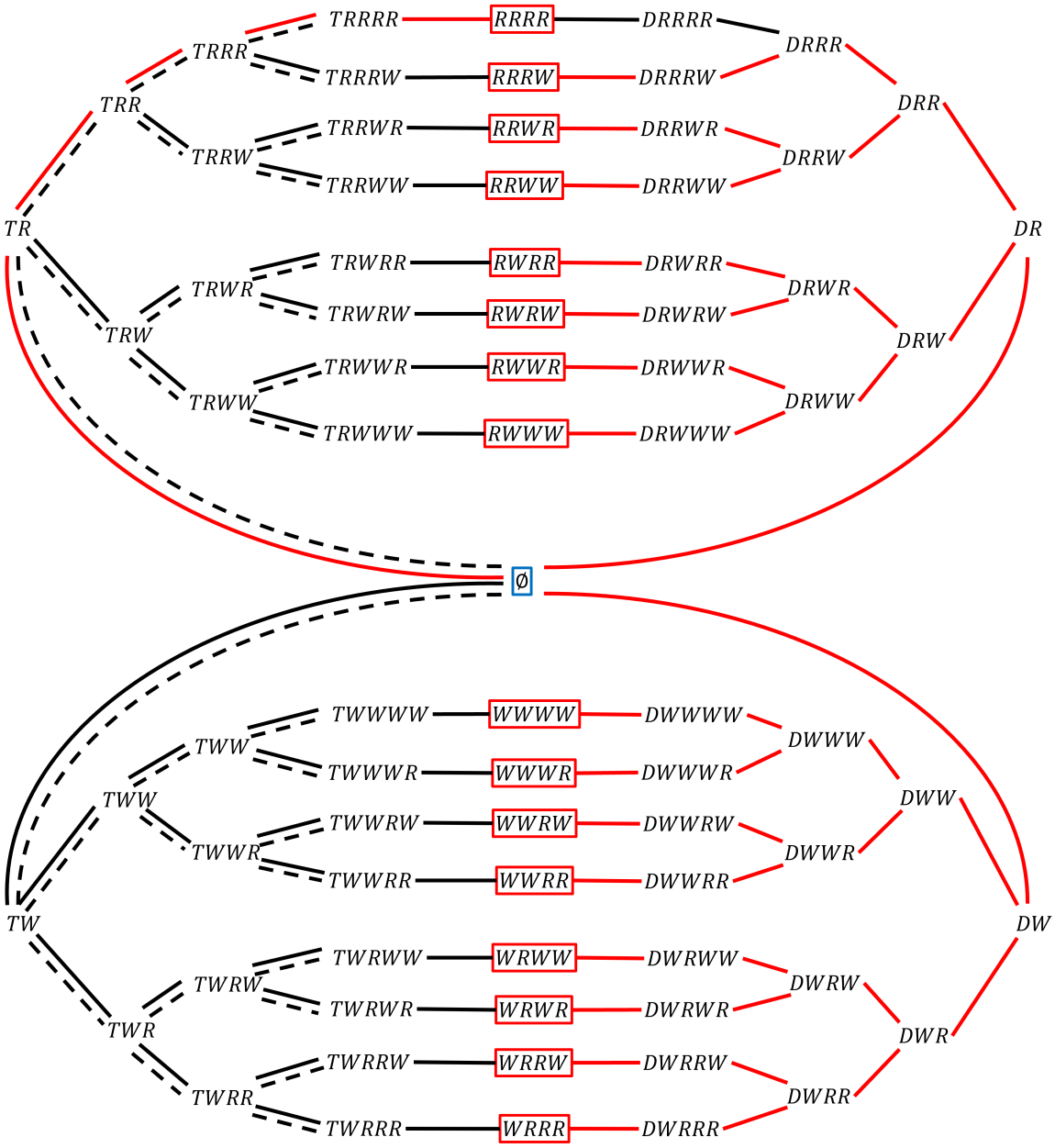


FIG. 11: Fast reactions (red) with rate ~ 1 for the naïve attempt to minimize the entropy of the product distribution.

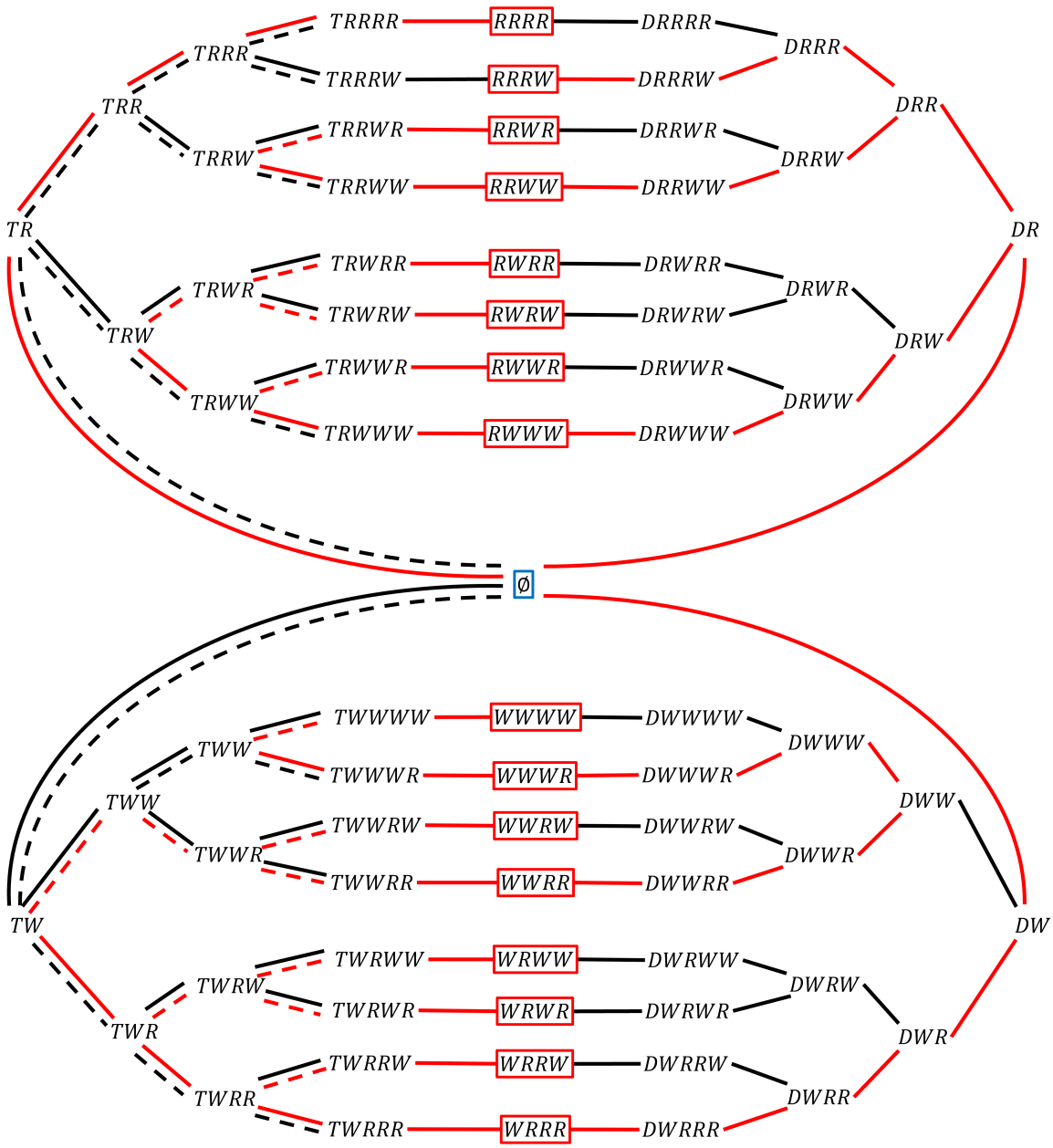


FIG. 12: Fast reactions (red) with rate ~ 1 for our best guess at how to minimize the entropy of the product distribution.