New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG IOP Institute of Physics

Edge-effects dominate copying thermodynamics for finite-length molecular oligomers

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
25 February 2021

REVISED
11 May 2021

ACCEPTED FOR PUBLICATION
20 May 2021

PUBLISHED
21 June 2021

PAPER

Edge-effects dominate copying thermodynamics for finite-length molecular oligomers

Jenny M Poulton and Thomas E Ouldridge*

Departmant of Bioengineering and Centre for Synthetic Biology, Imperial College London, London, SW7 2AZ, United Kingdom
* Author to whom any correspondence should be addressed.

E-mail: t.ouldridge@imperial.ac.uk

Keywords: biophysics, stochastic thermodynamics, information theory

Supplementary material for this article is available online

Abstract

A signature feature of living systems is their ability to produce copies of information-carrying molecular templates such as DNA. These copies are made by assembling a set of monomer molecules into a linear macromolecule with a sequence determined by the template. The copies produced have a finite length—they are often 'oligomers', or short polymers—and must eventually detach from their template. We explore the role of the resultant initiation and termination of the copy process in the thermodynamics of copying. By splitting the free-energy change of copy formation into informational and chemical terms, we show that, surprisingly, copy accuracy plays no direct role in the overall thermodynamics. Instead, finite-length templates function as highly-selective engines that interconvert chemical and information-based free energy stored in the environment; it is thermodynamically costly to produce outputs that are more similar to the oligomers in the environment than sequences obtained by randomly sampling monomers. In contrast to previous work that neglects separation, any excess free energy stored in correlations between copy and template sequences is lost when the copy fully detaches and mixes with the environment; these correlations therefore do not feature in the overall thermodynamics. Previously-derived constraints on copy accuracy therefore only manifest as kinetic barriers experienced while the copy is template attached; these barriers are easily surmounted by shorter oligomers.

Information transfer is the essence of the central dogma of molecular biology. The cell has sophisticated biomachinery that directs the copying of information from DNA into RNA, and then from RNA into proteins; complex molecules that perform varied roles within an organism [2]. The folding of proteins into their functional form is encoded directly into their sequence, meaning that accurate copying from DNA, via RNA, to proteins is vital for the functioning of an organism. In humans, there are around 20 000 functional proteins (peptide polymers) made from a library of 20 amino acids (monomer units), and the targeted assembly of these specific sequences by the cell is an astonishing accomplishment given the combinatorial space available to the system. The scale of this accomplishment can be highlighted by considering the difficulty of assembling such structures without templates.

Without a template, the cell would have to self-assemble proteins using hard-coded interactions between amino acid monomers targeting a specific structure as a free-energy minimum. This task is made difficult by the existence of many unintended, competing structures in cases where the same monomers are used in multiple different target structures [3]. Compounding this problem, the reliable self-assembly of even a single structure requires many distinct monomer types to encode the overall geometry. For example, a simple 4-armed cross, with arms of length two surrounding a central monomer, requires three distinct types of monomer to assemble reliably [4].

Targeting the reliable self-assembly of a single protein structure, containing hundreds of amino acid residues, would therefore be an extraordinary challenge with only 20 distinct monomers. Cells, however,

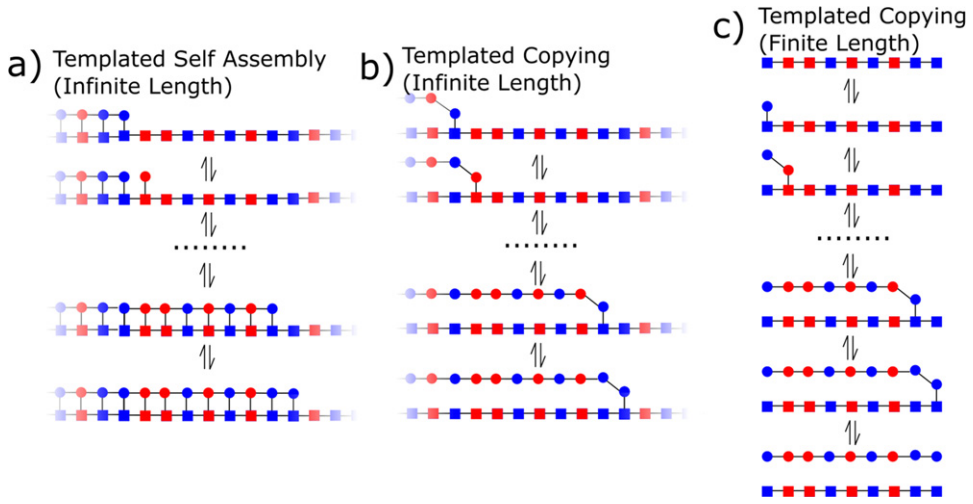

Figure 1. Distinct mechanisms for producing sequence-matched assemblies from a template. (a) A model for templated self-assembly on an infinitely long template, with a two-monomer alphabet for both template and product. Here favourable bonds retained between copy and template stabilise the low entropy product state. (b) A model of true templated copying on an infinitely-long template. The nascent copy detaches continuously from behind the leading edge of polymer growth, but the process never reaches the final monomer in the template and therefore the product remains bound by a single site. (c) A model of persistent copying on a finite-length oligomer template, including exchange of full length oligomers and single monomers with the environment.

need to target the assembly of tens of thousands of distinct proteins with the same 20 building blocks—it is simply impossible to hard-code this level of complexity into the interactions between the amino acids themselves [3].

Information transfer via templating solves this problem, because an arbitrary sequence of amino acids can be assembled in a way that does not depend on the interactions between the amino acids themselves. These interactions are only subsequently used to fold the protein into a specific secondary and tertiary structure, given the information in the sequence. It should be noted, however, that templating only solves this problem if the template is reusable. Without this crucial property, we would have merely pushed the problem back a layer; a new template would have to be assembled for each product created, recreating the problem solved by templating. Copy and template must therefore separate.

Numerous groups have worked on developing synthetic analogs of the natural polymer copying mechanisms. A major challenge is product inhibition, the tendency for copies to remain bound to their template, due to the cooperative interaction of the monomers in the copy with the template [5–7]. Cooperativity grows with length, meaning product inhibition is a more difficult problem to overcome for longer polymers. Consequently, progress in building synthetic copying systems where copy and template spontaneously separate—as in extant natural systems—has been limited. The most successful examples of copying rely on non-chemical or time-varying conditions to drive separation; the copy first forms on the template and then separates via mechanical scission [8, 9], heat [10], or a change in chemical conditions [11]. Due to the challenge of separating products from templates, synthetic examples of the biologically-relevant context in which copying is chemically-driven and autonomous have only involved dimers and trimers [12, 13].

This difficulty in recreating a fundamental biological phenomenon in a minimal synthetic context suggests a gap in our basic understanding. Indeed, most theoretical work has omitted the separation of copy and template [1, 14–20], instead studying templated self assembly, as illustrated in figure 1(a). In previous work [21], we considered a templated copying system including autonomous separation as shown in figure 1(b). Here, a growing polymer separates sequentially from the template as it grows, analogously to a nascent RNA or polypeptide chain strand during transcription or translation, respectively. The copy does not, therefore, build up a stronger and stronger attachment to the template. To put our results into context, we first summarise the fundamental thermodynamics of these two models as discussed in references [16, 21, 22].

For both templated self-assembly (omitting separation) and templated copying (including separation), producing a sequence that matches the template corresponds to producing a low entropy product. Throughout this work ‘low entropy product’ is used as a shorthand for ‘a product drawn from a low entropy distribution’. Considering a system with two monomer types, the reduction in sequence entropy (conditioned on the template sequence) for a perfect copy relative to an unbiased random sequence is

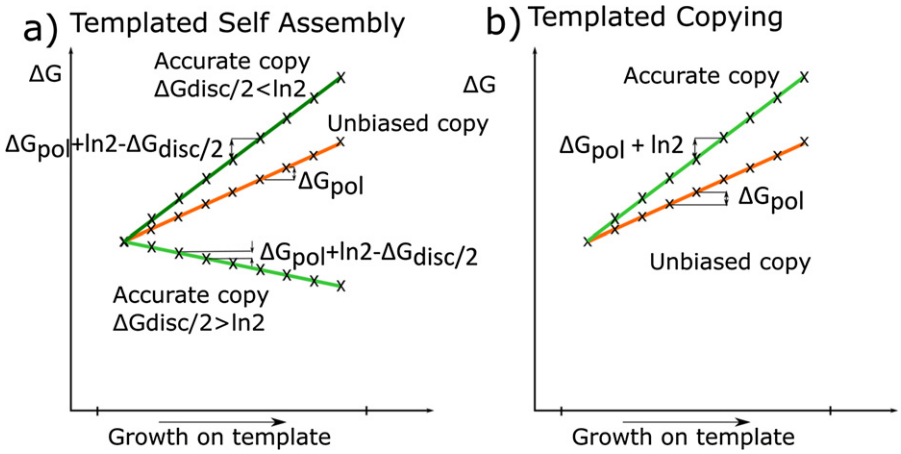

Figure 2. Thermodynamics of templated self-assembly and copying for infinite-length templates. Schematic free-energy profiles of (a) templated self-assembly on an infinite length template and (b) templated copying with continuous separation on an infinite-length template. $\Delta G_{\rm pol}$ is the total free energy change associated with adding a monomer to the growing polymer in an unbiased fashion and sets the slope of the orange line. In templated self-assembly, a perfectly accurate copy can be more favourable (have a more negative gradient) than an unbiased copy if the discrimination free energy difference is larger than the reduction in entropy due to creating an accurate copy-ln 2 per monomer (lighter green line) [1]. If however the discrimination free energy is smaller than ln 2 per monomer then the free energy of the accurate polymer will still be larger than the unbiased polymer (darker green line). For sufficiently large $\Delta G_{\rm disc}$ the green lines can be thermodynamically down hill even if the orange line is not. By contrast, for templated copying on an infinite-length template, accuracy *always* increases the gradient because the low entropy of the product cannot be compensated by sequence-specific bonds.

$\Delta H = \ln 2$ per monomer (throughout this article, we represent entropies in units of $k_{\rm B}$ and energies in units of $k_{\rm B}T$). Readers familiar with the idea of 'entropic growth' [16] will recall that a system that allows errors can grow under conditions when a system that does not allow errors would shrink. This observation is another way of expressing the same effect and the same consequence: the system that does not allow errors has lower entropy, and thus requires ln 2 per monomer more chemical free energy to be spent to lengthen the polymer chain than the system that creates an unbiased random sequence. Larger alphabet sizes would imply a larger entropy reduction—for example, an alphabet size of four as in transcription would give $\Delta H = \ln 4$.

To understand the role of this entropy difference, and the thermodynamic distinction between templated copying (including separation) and templated self-assembly (neglecting separation), consider the plots in figure 2. These plots give a representation of the non-equilibrium free-energy profile of copy formation as a visual summary of previous work from our group [21, 22] and others [16]. The plots are intended as schematics, but can also be rigorously interpreted as we outline in SI I (http://stacks.iop.org/NJP/23/063061/mmedia) [23]. For simplicity, all of these models have growth which is 'tightly coupled', without kinetic proofreading cycles [24], and in which the monomer interactions are symmetric (all complementary interactions are equivalent, as are all non-complementary interactions). We define $\Delta G_{\rm pol}$ as the average free energy change for increasing the length of a polymer with an unbiased, random sequence by one, which is the gradient of the free-energy profile for said sequence. These unbiased sequences will tend to grow if the slope is negative, $\Delta G_{\rm pol} < 0$. For a perfectly accurate copy, the reduction in entropy of $\Delta H = \ln 2$ per monomer relative to a random sequence will tend to make the gradient of the free-energy profile more positive by ln 2; the entropic cost thus tends to inhibit growth relative to an unbiased random sequence.

In a templated self assembly model that omits copy/template separation entirely (figure 1(a)), copy/template bonds are retained. This fact can overwhelm the entropic effect. Given free-energetically favourable interactions between matching sequences, correct matches are more strongly bound to the template than incorrect matches. Thus, the stronger copy/template interaction of a correctly-matched sequence bound to the template can more than compensate for the reduction in entropy required to create a perfectly accurate polymer. This behaviour is illustrated in figure 2(a) by the fact that the accurate sequence has a less positive gradient in its free energy profile than the unbiased sequence. Here $\Delta G_{\rm disc} > 0$ is the free-energy difference between a mismatched monomer binding to the template and a correctly-matched monomer binding to the template. A correctly-matched sequence bound to the template will have its free energy reduced by $\Delta G_{\rm disc}/2$ per monomer relative to an unbiased random sequence. For $\Delta G_{\rm disc}/2 > \ln 2$, the perfectly accurate polymer bound to the template is thermodynamically more favourable than an unbiased sequence because the discrimination free energy has overcome the extra entropic cost [1]. Thus

the perfectly accurate polymer has a more negative slope in the free-energy profile and grows at a less negative ΔG_{pol}. In principle, ΔG_{disc} could be arbitrarily large, making the perfectly-matched sequence arbitrarily thermodynamically favoured relative to alternatives. Polymers produced by templated self-assembly (neglecting separation) can thus be arbitrarily accurate in thermodynamic equilibrium.

Explicitly considering the disruption of copy-template bonds, as in figure 1(b), changes this analysis. [21, 22] Since all copy-template bonds are transient, there is no long-term thermodynamic benefit to incorporating a correctly-matched monomer into the copy. Consequently, the entropic cost of creating an accurate copy that separates from the template cannot be overcome by ΔG_{disc} terms, and the free-energy profile of an accurate copy always has a steeper slope than a random sequence (figure 2(b)). Unlike templated self-assembly, therefore, the thermodynamic cost of accuracy cannot be compensated for by copy-template bonds: instead, a more negative ΔG_{pol} (a stronger thermodynamic drive towards polymerisation) is required to grow an accurate copy. In addition, the higher free energy of the accurate sequence means that, unlike in templated self-assembly, accurate copying including separation necessitates generating a far-from-equilibrium product [21].

Mechanistically, producing a specific non-equilibrium state is a very different challenge from allowing the system to relax to a stable equilibrium. Moreover, from the thermodynamic perspective, the template is playing a distinct role. It acts as an engine that transduces excess free energy stored in the input monomers into excess free energy stored in the low entropy product sequence, rather than as a reactant whose sequence directly contributes to the free energy of products. It was observed in reference [21], that, for finite binding free energies, this transduction could never be 100% efficient. Producing an accurate copy was therefore seen to be necessarily thermodynamically irreversible.

These previously-obtained results, both for templated self-assembly [1, 14–20] and for a polymer that continuously separates from its template [21], were derived for infinite-length polymers. The tip of the growing copy is assumed to reach a steady velocity along the template, and the identity of monomers measured relative to this tip reach a stationary distribution. Initiation and termination of the polymerization process were ignored, and even in reference [21] the copy remains attached by a single bond—complete detachment was neglected.

Although this approximation might be reasonable for some long biopolymers *in vivo*, it is a poor approximation for the copying of shorter oligomers and dimers. Given that synthetic systems are currently limited to short oligomers, and that early life is likely to have created short oligomers before the origin of complex enzyme-based copying machinery, it is worth studying initiation and termination in more detail. Equally, 'charged' tRNA—hybrid molecules of one tRNA codon and its matching amino acid that are necessary for the process of RNA translation into proteins—are dimerised via a specific template enzyme called a synthetase [25]. This process is essentially the copying of a dimer template, and so the creation of short copies is highly relevant even to extant biology.

In this article we probe the consequences of the 'edge-effects' of initial attachment and final detachment on the copying of oligomer sequences. We first consider the free-energy change for the production of a single finite-length copy under constant external conditions, separating it into chemical and informational terms. Using dimerisation as an example, we show that the overall thermodynamic constraints on information transfer are fundamentally altered relative to infinite-length polymers: in general there is no entropic cost to accurately reproducing the template sequence, in direct contrast to previous work on infinite-length templates [1, 14–21]. As in reference [21], the template acts as an information engine that does not directly contribute to the relative free energies of products, but here its role is to selectively couple to a subset of a large number of out-of-equilibrium molecular reservoirs; the relationship between the copy sequence and these reservoirs sets the overall thermodynamics. Although these results hold for products of arbitrary finite length, we nonetheless observe a gradual cross-over to the previously predicted constraints on accuracy [21] when studying a particular dynamical model of copying with longer oligomers (figure 2(c)). The thermodynamic constraints on accuracy in the infinite-length limit [21] instead become kinetic barriers for finite-length templates.

1. Model of oligomerisation

Figure 3 shows a dimerisation system, a prototype for a broader class of oligomerisation systems studied here. A solvated template dimer carries information in its sequence of monomer units; in this case, the sequence is 1,1. The template is coupled to large baths of monomers and oligomers (in this case dimers) of a distinct second type of molecule, like a DNA template in a bath of RNA nucleotides and oligomers. This second type of molecule also comes in multiple varieties—in this case two—and can interact with the template in a sequence-specific way. In figure 3, we propose a particular thermodynamically-consistent

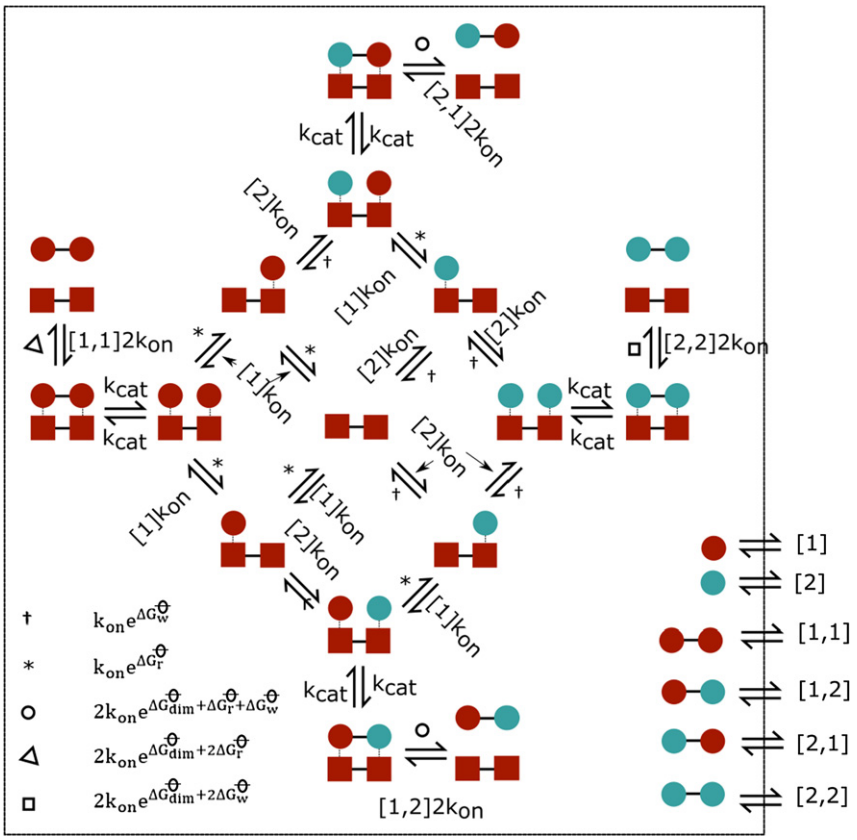

Figure 3. Minimal model of oligomerisation illustrated with dimers. A template (squares) interacts with baths of monomers and dimers of a second class of molecules (circles) which is present as type '1' or type '2'. Monomers can bind to the template and dimerize, while dimers binding to the template can be destroyed or interconverted. The standard dimerisation free-energy is $\Delta G^{\ominus}_{\mathrm{dim}}$ for all sequences, but matching and non-matching varieties of monomer bind to the template with strengths $\Delta G^{\ominus}_{r}$ and $\Delta G^{\ominus}_{w}$, respectively, allowing selectivity. Rate constants k_{cat} and k_{on} define the dynamics.

model for dimer production (for simplicity, we do not allow for kinetic proofreading cycles [24] in our analysis). Dimers from the baths can also be broken down into their component monomers by the template.

The parameters in figure 3 can be divided into two classes; those that are properties of the copy species in isolation, and those that relate to the interaction of the copy species with the template. With regard to the former, we define $\Delta G^{\ominus}_{\mathrm{dim}}$ as the free-energy change of dimerisation at the reference concentration for all sequences. We assume $\Delta G^{\ominus}_{\mathrm{dim}}$ is sequence-independent, since arbitrary copy sequences must be producible given the right template. The concentrations of the two monomer types are $[1]$ and $[2]$ (defined with respect to a standard reference concentration), and the four polymer types are $[1,1]$, $[1,2]$, $[2,1]$ and $[2,2]$.

The properties of the copy species alone set the thermodynamic constraints on the system that we will derive in section 2.1. The details of the interaction between copy molecules and template will determine how a specific example system behaves with respect to these constraints. For the dimerisation example in figure 3, $\Delta G^{\ominus}_{r}$ is the standard free-energy change of a matching monomer binding to the template; $\Delta G^{\ominus}_{w}$, is the equivalent for a mismatch. The difference between the two, $\Delta G_{\mathrm{disc}} = \Delta G^{\ominus}_{w} - \Delta G^{\ominus}_{r}$, is the discrimination free energy. Parameters k_{cat} and k_{on} set the absolute values of transition rates between states.

In the model of figure 3, we assume that any free energy released by dimerisation is used to destabilise the bonds between dimer and template. The state with two monomers and the states with a dimer bound to the template therefore have equal free energy, with k_{cat} the underlying rate of the formation or breaking of the backbone bond. Such a free-energy landscape has been proposed to be optimal for minimising product inhibition [26]. Here, matching and non-matching monomers bind to the template with the same rate constant k_{on}, but since $\Delta G^{\ominus}_{r} < \Delta G^{\ominus}_{w}$, mismatches detach faster. Since we use an unbiased $m(s)$ in all case studies, the symmetry of the problem gives identical physics for all template sequences; we shall use 1,1 for clarity.

2. Results

2.1. Thermodynamics of oligomerisation

For a general dimerisation model, the total free-energy change of the baths upon creating a single dimer of sequence $\mathbf{s}$ is $\Delta G(\mathbf{s}) = \Delta G^{\ominus}_{\mathrm{dim}} + (\ln[\mathbf{s}] - \ln[s_1][s_2])$; here $\mathbf{s}$ is the arbitrary sequence s_1, s_2. $\Delta G^{\ominus}_{\mathrm{dim}}$ is the sequence-independent free-energy for forming a backbone bond under standard conditions, which includes the standard chemical potentials of the monomers and oligomer. $(\ln[\mathbf{s}] - \ln[s_1][s_2])$ is the contribution to the free-energy change provided by the monomer and oligomer baths. For a given sequence $\mathbf{s}$, there is no net production or destruction of dimers at $\Delta G(\mathbf{s}) = 0$.

Since the template itself acts catalytically [22], its free-energy is unchanged and thus does not contribute to $\Delta G(\mathbf{s})$. Since we do not allow for kinetic proofreading cycles [24] in our analysis, the process is tightly coupled and there is no unknown and variable number of fuel-consuming futile cycles. The free energy change of the dimerisation process as a whole is is therefore unambiguously defined as $\Delta G(\mathbf{s})$, regardless of the details of the copying mechanism.

This free-energy change of dimerisation can be generalised straight-forwardly to oligomers of length $|\mathbf{s}|$,

$$\Delta G(\mathbf{s}) = (|\mathbf{s}| - 1)\Delta G^{\ominus}_{\mathrm{dim}} + \left(\ln[\mathbf{s}] - \ln \prod_{i=1}^{|\mathbf{s}|} [s_i]\right). \tag{1}$$

where $\Delta G^{\ominus}_{\mathrm{dim}}$ is now the standard free energy change for adding any monomer to the end of a copy oligomer in solution.

We now move to defining the average rate of change of free energy for a system of a template interacting with multiple monomer and oligomer types assuming a stationary state for the template, resulting in a steady flux of production or destruction for each sequence. Let $J(\mathbf{s})$ be the expected net rate at which sequence $\mathbf{s}$ is produced by the system for a given constituent monomer and oligomer concentrations and $\Delta G^{\ominus}_{\mathrm{dim}}$. The average rate of change of free-energy is then $\Delta \dot{G} = \sum_{\mathbf{s}} J(\mathbf{s})\Delta G(\mathbf{s})$. We define the normalised flux $q(\mathbf{s}) = J(\mathbf{s})/J_{\mathrm{tot}}$. When $q(\mathbf{s}) \geqslant 0$ for all $\mathbf{s}$, $q(\mathbf{s})$ can be treated as the probability of picking sequence $\mathbf{s}$ from the flow of products being created by the template. When this is not the case then it must be understood only as a normalised flux. We further define the following probability distributions: $p(\mathbf{s}) = [\mathbf{s}]/[S_{\mathrm{tot}}]$, the probability of picking an oligomer of sequence $\mathbf{s}$ from the oligomer baths with total concentration $[S_{\mathrm{tot}}]$; $m(s) = [s]/[M_{\mathrm{tot}}]$, the probability of picking a monomer of type s from the monomers with total concentration of $[M_{\mathrm{tot}}]$; and $t(\mathbf{s}) = \prod_i m(s_i)$, which corresponds to the probability of the sequence $\mathbf{s}$ occurring by selecting monomers randomly from the monomer pools. In these terms,

$$\Delta \dot{G} = J_{\mathrm{tot}} \left(\sum_{\mathbf{s}} q(\mathbf{s}) \left(|\mathbf{s}| - 1\right) \Delta G^{\ominus}_{\mathrm{dim}} \right) + J_{\mathrm{tot}} \left(\sum_{\mathbf{s}} q(\mathbf{s}) \ln \frac{p(\mathbf{s})}{t(\mathbf{s})} + \sum_{\mathbf{s}} q(\mathbf{s}) \ln \frac{[S_{\mathrm{tot}}]}{[M_{\mathrm{tot}}]^{|\mathbf{s}|}} \right).$$

In general the equilibrium point of this system is when $\Delta \dot{G} = 0$, requiring $J(\mathbf{s}) = 0$ for all $\mathbf{s}$. This expression can be re-written as

$$\Delta \dot{G} = J_{\mathrm{tot}} \Delta G_{\mathrm{chem}} + J_{\mathrm{tot}} \Delta G_{\mathrm{inf}}, \tag{2}$$

with

$$\begin{aligned} \Delta G_{\mathrm{chem}} &= (|\mathbf{s}| - 1)\Delta G^{\ominus}_{\mathrm{dim}} + \ln[S_{\mathrm{tot}}]/[M_{\mathrm{tot}}]^{|\mathbf{s}|}, \\ \Delta G_{\mathrm{inf}} &= \sum_{\mathbf{s}} q(\mathbf{s}) \ln \frac{p(\mathbf{s})}{t(\mathbf{s})}, \end{aligned} \tag{3}$$

assuming for simplicity that all oligomers are of the same length. The first term in equation (2) is the average chemical free-energy change of oligomerisation ignoring sequence, multiplied by the net rate of oligomer production. The second term is information-theoretic: for non-negative net production of all oligomers $J(\mathbf{s}) = q(\mathbf{s})J_{\mathrm{tot}} \geqslant 0$ for all $\mathbf{s}$, $\Delta G_{\mathrm{inf}} = D(q||t) - D(q||p)$, where $D(q||p) = \sum_{\mathbf{s}} q(\mathbf{s}) \log \frac{q(\mathbf{s})}{p(\mathbf{s})}$ is the Kullback–Leibler divergence between $q(\mathbf{s})$ and $p(\mathbf{s})$. ΔG_{inf} reflects the sequence statistics of monomer and oligomer baths, and the sequence-dependence of net oligomer production. This splitting into chemical and informational terms holds for arbitrary oligomer lengths and sequence alphabets, and is the first result of this paper.

2.2. Is there necessarily a thermodynamic cost to accuracy?

In all previous work on infinite-length templates there is a thermodynamic cost to creating an accurate, low entropy copy. In templated self-assembly systems where separation is omitted altogether [1, 14–20], low

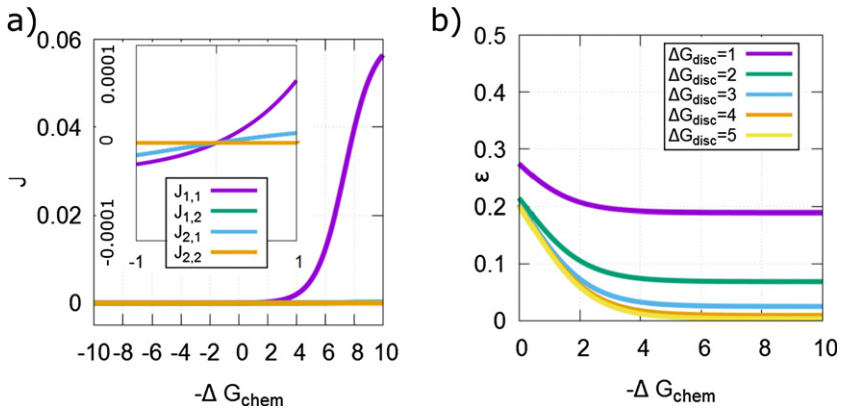

Figure 4. A dimer copying model shows finite accuracy in the equilibrium limit. (a) The net production rate $J(\mathbf{s})$ against chemical driving $-\Delta G_{\rm chem}$ for each of the four dimers 1,1, 1,2, 2,1 and 2,2 relative to a template of sequence 1,1, with $\Delta G_{\rm disc} = \Delta G_w^{\ominus} - \Delta G_r^{\ominus} = 10$, $\Delta G_r^{\ominus} + \Delta G_w^{\ominus} = 0$ and unbiased monomer and oligomer baths. $\Delta G_{\rm chem}$ is adjusted by changing the standard dimerisation free energy $\Delta G_{\rm dim}^{\ominus}$. All $J(\mathbf{s})$ pass through zero at the equilibrium point $\Delta G_{\rm chem} = 0$ but quickly separate when $\Delta G_{\rm chem} \neq 0$ (inset). $J(1,2)$ and $J(2,1)$ overlap. At larger driving, accurate copies are preferentially produced. (b) Error fraction at which incorrect type 2 monomers are incorporated into dimers, ϵ, against $\Delta G_{\rm chem}$ for various $\Delta G_{\rm disc}$ and $\Delta G_r + \Delta G_w = 0$. The system has finite accuracy, $\epsilon \neq 0.5$, as $\Delta G_{\rm chem} \to 0^-$, with the accuracy dependent on $\Delta G_{\rm disc}$.

entropy sequences can be compensated by favourable bonds between copy and template. In our previous work on templated copying—which incorporates continuous separation from behind the leading edge of oligomer growth but omits the final separation from the template [21]—this entropic cost has to be paid for directly by the free energy of polymerisation.

We now use our exemplar system, shown in figure 3, to explore the question of the thermodynamic cost of accuracy in the context of a system with finite-length templates, before drawing general conclusions by relating the results to equation (3). It is, in principle, possible to fix the monomer and polymer baths at arbitrary concentration. Throughout this work, we assume constant bath concentrations, and calculate the flux $J(\mathbf{s})$ via $q(\mathbf{s})$ in steady state by analysing the Markov process corresponding to the states of a single isolated template [23]. Let us first consider a template of sequence 1,1 coupled to baths where all oligomers have the same concentration $[S_{\rm tot}]/4$, and all monomers have the same concentration $[M_{\rm tot}]/2$. Without loss of generality we may choose our standard concentration so that $[S_{\rm tot}]/[M_{\rm tot}]^2 = 1$, and thus $\Delta G_{\rm chem} = \Delta G_{\rm dim}^{\ominus}$.

In figure 4(a) we plot the net production rate of each dimer as a function of $\Delta G_{\rm chem}$ at set $\Delta G_{\rm disc} = \Delta G_r^{\ominus} - \Delta G_w^{\ominus}$. In general, the equilibrium point of our dimerisation system is when $\Delta \dot{G} = 0$ and where $J(\mathbf{s}) = 0$ for all $\mathbf{s}$. In this case, this point occurs when $\Delta G_{\rm chem} = 0$, with net creation for all dimers if $\Delta G_{\rm chem} < 0$ and net destruction if $\Delta G_{\rm chem} > 0$. In figure 4(b) we plot $\epsilon = \left(J(2,2) + \frac{1}{2}(J(1,2) + J(2,1))\right)/J_{\rm tot}$, the proportional rate at which incorrect monomers are incorporated into dimers. It is noticeable that while at exactly $\Delta G_{\rm chem} = 0$, ϵ is undefined, as $\Delta G_{\rm chem} \to 0^-$ we obtain $\epsilon < 0.5$, implying non-zero accuracy.

Figure 5 shows that when the average template binding free-energy of right and wrong monomers is increased, at fixed $\Delta G_{\rm disc} = \Delta G_w^{\ominus} - \Delta^{\ominus} G_r$, the error remains low as the system tends towards equilibrium ($\Delta G_{\rm chem} \to 0^-$ for unbiased $p(\mathbf{s})$, $t(\mathbf{s})$ and $[S_{\rm tot}]/[M_{\rm tot}] = 1$). Indeed, $\epsilon \to 0$ (perfect accuracy) as $\Delta G_{\rm disc} \to \infty$ and $\Delta G_{\rm chem} \to 0^-$. The error at $\Delta G_{\rm chem} = 0$ is still undefined, but the fluxes separate quickly after this point (inset figure 5(a)) to keep the error low. Due to the unstable bonds between copy and template, the system has a low flux for $\Delta G_{\rm chem} < 0$.

To summarize, the minimal value of $-\Delta G_{\rm chem}$ required for growth gives non-zero accuracy, and reversible processes can create a low entropy dimer sequence distribution at finite discrimination free-energy $\Delta G_{\rm disc}$. However, interactions between template and product do not persist; $\Delta G_{\rm disc}$ does not feature in the overall thermodynamics, and cannot compensate for a low entropy state in equilibrium, as in templated self-assembly [1, 14, 15, 17–20].

To understand this apparent absence of a thermodynamic cost to accuracy, consider the general result in equation (3). For the results plotted in figure 4, $p(\mathbf{s})$ and $t(\mathbf{s})$ are unbiased and equal, and thus $\Delta G_{\rm inf} = 0$ for any $q(\mathbf{s})$—even if only accurate copies are produced. More generally, whenever the surrounding oligomers have the same sequence distribution that would be found by randomly selecting monomers from the monomer baths (i.e., $p(\mathbf{s}) = t(\mathbf{s})$), there is no extra thermodynamic cost to producing sequences of arbitrary accuracy. We emphasise the surprising result that arbitrary accuracy has no extra cost *even if* the monomer distribution is unrelated to the template sequence.

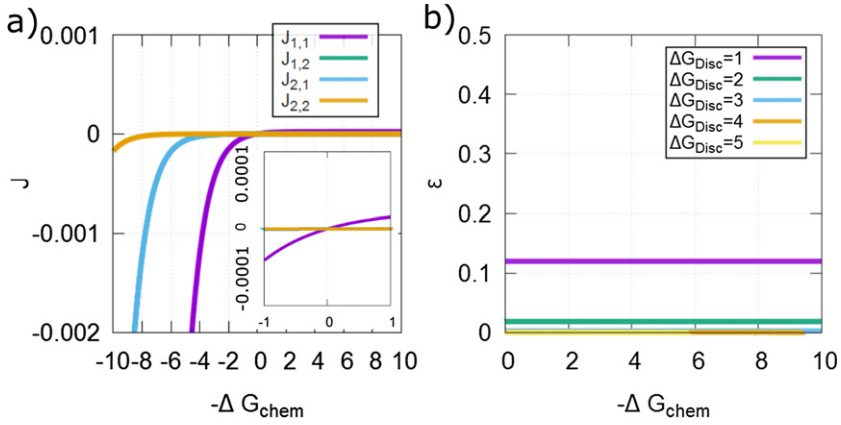

Figure 5. The dimer copying model can show arbitrarily high accuracy in the equilibrium limit. (a) We plot the net production rate J_s against chemical driving $-\Delta G_{\text{chem}}$ for each of the four dimers 1,1, 1,2, 2,1 and 2,2 relative to a template of sequence 1,1 and with $\Delta G_{\text{disc}} = \Delta G^{\ominus}_{w} - \Delta G^{\ominus}_{r} = 2$, but $\frac{\Delta G^{\ominus}_{r} + \Delta G^{\ominus}_{w}}{2} = 5$. All oligomers have the same (fixed) concentration $[S_{\text{tot}}]/4$, and all monomers have the same (fixed) concentration $[M_{\text{tot}}]/2$. All J_s pass through zero at the equilibrium point $\Delta G_{\text{chem}} = 0$ but separate quickly (inset) at non-zero driving. Here accurate copies are preferentially produced but all fluxes are very low. (b) Error fraction at which incorrect type 2 monomers are incorporated into dimers, ϵ, against $-\Delta G_{\text{chem}}$ for various ΔG_{disc} with $\frac{\Delta G^{\ominus}_{r} + \Delta G^{\ominus}_{w}}{2} = 5$. The system tends towards perfect accuracy, $\epsilon \to 0$, as $\Delta G_{\text{disc}} \to \infty$, even as the system tends to the equilibrium point $\Delta G_{\text{chem}} \to 0^{-}$.

ΔG_{inf} is generally non-zero, however, for systems with $p(\mathbf{s}) \neq t(\mathbf{s})$, i.e., where the monomer and oligomer baths have different distributions. ΔG_{inf} is positive if the system produces sequences that are common in the oligomer bath $p(\mathbf{s})$ and rare in the monomer bath $t(\mathbf{s})$. The alternative representation $\Delta G_{\text{inf}} = D(q\|t) - D(q\|p)$ makes this fact particularly clear; the probability distribution of the creation fluxes $q(\mathbf{s})$ being similar to the monomer distribution $t(\mathbf{s})$ makes the first term less positive and $q(\mathbf{s})$ being unlike the oligomer bath $p(\mathbf{s})$ makes the second term more negative. Accuracy is therefore not directly constrained by thermodynamics in a general description of the full process of oligomer copying. Instead, there is a thermodynamic cost to producing sequence distributions $q(\mathbf{s})$ that are closer to the oligomer sequences in the environment than a distribution of sequences obtained by randomly sampling monomers from the environment. This argument is the second main result of this paper.

3. Template copying as an inherently non-equilibrium information engine

Unlike previous models of templated self-assembly [1, 14, 15, 17–20], in this context the template act cyclically. The template can therefore be thought of as an engine in the conventional thermodynamic sense, like a Carnot engine. Here, the engine operates between the monomer and oligomer reservoirs. Its thermodynamics is set by the relationship between its input from, and its output to, these reservoirs relative to the chemical potential of those reservoirs. This is the physics encapsulated by equation (2).

Unusually, the system studied here can couple to so many reservoirs that it is generally impossible to find an equilibrium point where $\Delta\dot{G} = 0$, and where $J(\mathbf{s}) = 0$ for all $\mathbf{s}$. To illustrate, consider a set of systems with $p(\mathbf{s}) \neq t(\mathbf{s})$, i.e., where the monomer and oligomer baths are unbalanced. Here, there is no point at which all fluxes are zero because there is no way to simultaneously bring all baths into equilibrium with each other simply by changing ΔG_{chem} via $\Delta G^{\ominus}_{\text{dim}}$. There is instead a range of ΔG_{chem} over which $J_{\text{tot}} = \sum_{\mathbf{s}} J(\mathbf{s}) = 0$ could occur (with some $J(\mathbf{s}) > 0$ and some $J(\mathbf{s}) < 0$), depending on which sequences best couple to the template. The most positive possible ΔG_{chem} at which $J_{\text{tot}} = 0$ occurs is observed when a system specifically produces the sequence $\mathbf{s}_{\text{min}}$, where $\mathbf{s}_{\text{min}}$ minimises $t(\mathbf{s})/p(\mathbf{s})$. The most negative is when the system specifically produces the sequence $\mathbf{s}_{\text{max}}$, where $\mathbf{s}_{\text{max}}$ maximises $t(\mathbf{s})/p(\mathbf{s})$. $\mathbf{s}_{\text{min}}$ is intuitively the sequence most like the monomer baths and least like the polymer baths and $\mathbf{s}_{\text{max}}$ is the opposite. In figure 6(a), we vary ΔG_{disc} for a system heavily thermodynamically biased towards creating accurate copies by the baths. When $\Delta G_{\text{disc}} > 0$, and the system is also kinetically biased towards creating accurate copies and $J_{\text{tot}} = 0$ for a more positive ΔG_{chem} than if $\Delta G_{\text{disc}} < 0$. Copying accurately can thus either make production of oligomers thermodynamically easier or harder, depending on the environment. This fact is true even for an unbiased pool of monomers.

We define $q_{\text{min}}(\mathbf{s})$ as the $q(\mathbf{s})$ that results in the most negative value of ΔG_{inf}, and $q_{\text{max}}(\mathbf{s})$ which maximises ΔG_{inf}. When the probability distribution of fluxes $q(\mathbf{s})$ is close to $q_{\text{min}}(\mathbf{s})$ it is possible for a negative ΔG_{inf} to overcome a positive ΔG_{chem}. Equally, a more negative ΔG_{chem} allows for a $q(\mathbf{s}) \approx q_{\text{max}}(\mathbf{s})$

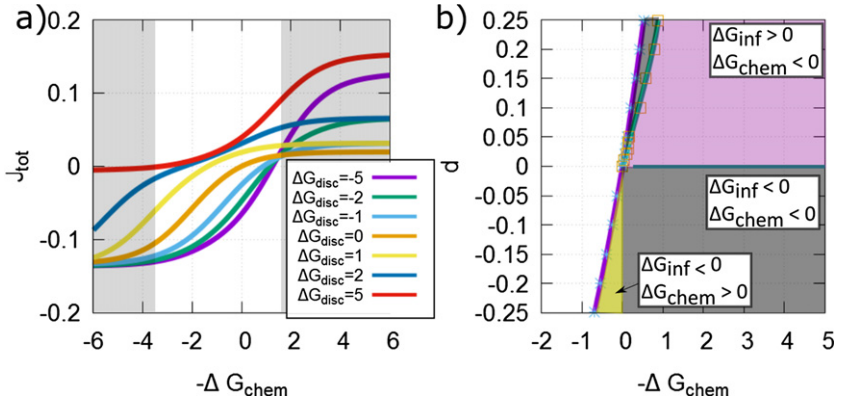

Figure 6. Oligomer copying as an information engine with no equilibrium point. (a) Total flux J_{tot} against driving $-\Delta G_{\text{chem}}$ for a range of discrimination free energies $\Delta G_{\text{disc}} = \Delta G_w^{\ominus} - \Delta G_r^{\ominus}$, with $[1] = [2] = 0.1$, $[1,1] = [1,2] = [2,1] = 0.001$ and $[2,2] = 0.1$. ΔG_{disc} is varied with $\Delta G_r^{\ominus} + \Delta G_w^{\ominus} = 0$ fixed. The point $J_{\text{tot}} = 0$ at which there is no net dimerisation varies within the allowed white range despite the fact that the overall dimerisation free-energy is independent of $-\Delta G_{\text{disc}}$. Specificity for $\mathbf{s}_{\min} = 1, 1$ makes growth easier and pushes $J_{\text{tot}} = 0$ to the lower limit, and specificity for $\mathbf{s}_{\max} = 2, 2$ has the opposite effect. (b) Phase plot of the information engine. Here the leftmost purple boundary is the transition from J_{tot} negative to positive. Here we fix $\Delta G_{\text{disc}} = 5$, $[S_{\text{tot}}] = 1$, $[M_{\text{tot}}] = 1$, $t(s) = 0.25$ for all s and vary ΔG_{chem}. We further vary $p(s) = 0.25 + d, 0.25, 0.25, 0.25 - d$ by varying d. There is a regime in which chemical work is used to specifically produce sequences of high free-energy and a regime in which specific production of low free-energy sequences is used to drive oligomerisation against a chemical load.

with positive ΔG_{inf}. From this perspective, thinking about the monomer and oligomer reservoirs as a single collective environment, we can describe the template as an engine that trades chemical and information-based free energy in the environment against each other [27, 28].

The second law implies that the rate of change of free energy $\Delta\dot{G}$ is non-positive. Thus, from equation (2), there are three possible regimes for this information engine, illustrated in figure 6(b). If $\Delta G_{\text{chem}} < 0$ and $\Delta G_{\text{inf}} > 0$ then the system channels chemical work through a specific copying mechanism to store free energy in a distribution of outputs closer to the oligomer bath $p(\mathbf{s})$ than the monomer bath $t(\mathbf{s})$, with an efficiency $\eta = \frac{\Delta G_{\text{inf}}}{-\Delta G_{\text{chem}}} \leqslant 1$. In our case, η reaches a maximum of $\sim$0.3 when $p(\mathbf{s})$ is heavily biased towards accurate copies of the template and ΔG_{chem} is small and negative. In the case where $\Delta G_{\text{chem}} > 0$ and $\Delta G_{\text{inf}} < 0$, the system generates outputs closer to the monomer baths $t(\mathbf{s})$ than the oligomer baths $p(\mathbf{s})$, expending information to compensate for an unfavourable chemical work term. Here the efficiency $\eta = \frac{\Delta G_{\text{chem}}}{-\Delta G_{\text{inf}}} \leqslant 1$ reaches a maximum of 0.15 when $p(\mathbf{s})$ is heavily biased against accurate copies of the template and ΔG_{chem} is small and positive. The final case, in which both $\Delta G_{\text{chem}} \leqslant 0$ and $\Delta G_{\text{inf}} \leqslant 0$, is less interesting as the system both spends chemical free energy and generates outputs close to $q(\mathbf{s}) = q_{\min}(\mathbf{s})$.

4. Kinetic convergence on thermodynamic constraints for infinite-length polymers

The results of this paper—including the absence of an intrinsic thermodynamic cost to accuracy—apply to polymers of arbitrary finite length. So are the thermodynamic constraints on copy accuracy derived in reference [21], discussed in the introduction and illustrated in figure 2, irrelevant when initiation and eventual termination are explicitly considered, no matter how long the template and copy are?

Although initiation and termination fundamentally change the overall thermodynamics of copying, the bulk of the process in which the copy oligomer grows on the template is identical to the infinite-length case. To illustrate this comparison between the results of section 2.2 and those of reference [21], we return to the free-energy profiles of figure 2, reproducing the plot for templated copying in figure 7(a). In figure 7(b), we illustrate schematically the effect of considering initiation and termination via a model such as that in figure 1(c).

The identical gradients in figure 7(a) and (b) illustrate the principle that the intermediate steps of growth on the template are equivalent whether the template is infinite, or finite with explicit initiation and detachment steps. Those initiation and detachment steps, however, have a free-energy change that is different from the free-energy change of growth on the template. Initiation and detachment include no polymerisation, and final detachment of the copy from the template involves a large entropy increase due to the copy no longer being spatially confined to the template. These 'edge effects' ensure that the overall free-energy change is given by equation (2), rather than simply the copy length multiplied by the gradient of the profile when template-attached.

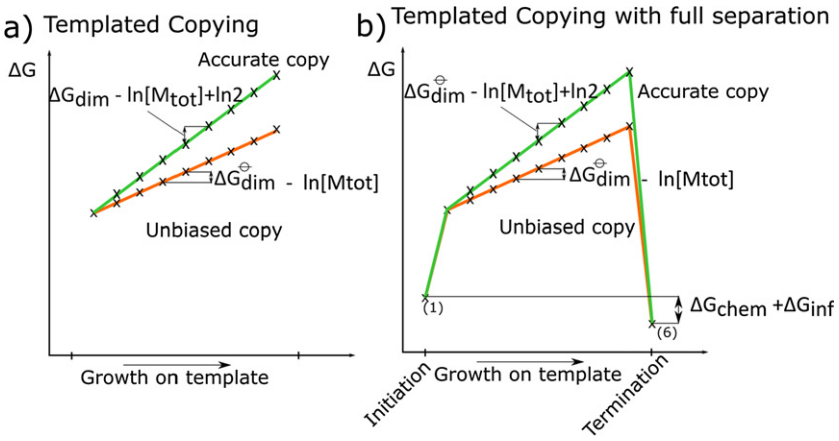

Figure 7. Schematic free-energy profiles for (a) a templated copying process on an infinite-length template and (b) a finite length template with explicit initiation and termination, in a system with uniform $p(\mathbf{s})$ and $t(\mathbf{s})$. The gradient of the free-energy profile during the bulk of the copying process is set by the standard free-energy of dimerisation, the monomer concentration and the accuracy of the copy. For a system with two monomer types, an accurate copy requires an extra ln 2 to add a monomer relative to an unbiased copy. However in the case of a finite length template, the initiation and termination steps involve physically distinct processes, resulting in steps in the profile that do not match the gradient of growth on the template. These steps can be large, and ensure that the overall process has the correct free-energy change.

Indeed, initiation and termination provide a theoretically unlimited adjustment to the overall ΔG, since the overall free-energy change depends on $\ln[S_{\text{tot}}]$ and $D(q\|p)$. Both terms can be arbitrarily large but do not contribute to the gradient of the free energy profile while the copy is template-bound. An arbitrarily unfavourable polymerisation process on the template (a free energy profile with a steep positive slope) can thus be made favourable overall with the right monomer and oligomer environment.

Importantly, however, if template-attached growth is unfavourable (i.e., the free-energy profile is uphill during the copy process), it will be kinetically suppressed by a large free-energy barrier even if oligomer production is favourable overall. Here, barrier height grows proportionally to oligomer length, suggesting that the *thermodynamic* constraints derived for infinite-length polymers should become increasingly prohibitive *kinetic* effects for sufficiently long oligomers. To probe this hypothesis we consider a kinetic model for the growth and destruction of oligomers of arbitrary fixed length, by extending the model of reference [21] so that it includes initiation and termination (figure 1(c)).

4.1. Model of copy production for oligomers of length $|s| > 2$

The model of reference [21] considers a single template consisting of a series of monomers $\mathbf{n} = n_1, \ldots, n_L$, with $L \to \infty$. Growing on the template is a single copy oligomer $\mathbf{s} = s_1, \ldots, s_l$, $(l \leqslant L)$. Inspired by transcription and translation, the model describes a copy that detaches from the template as it grows. Transitions in the model are whole steps in which a single monomer is added or removed from the copy's leading edge, potentially encompassing many individual chemical sub-steps (for ease of discussion we assume that the overall dynamics is well modelled by a single instantaneous step, but that is not essential for our conclusions). As illustrated in figure 1(c), after each step there is only a single inter-polymer bond at position l, between s_l and n_l. As a new monomer joins the copy at position $l+1$, the bond at position l is broken.

We augment this existing model by restricting the template to a length L, and considering the possibility of an empty template as shown in figure 1(c). Monomers can bind to or detach from the first site of an otherwise empty template, and full-length oligomers can bind to and detach from the final site of a template. We assume that at most one copy is bound to the template at any time.

As in the dimerisation model considered earlier in the text, we shall consider a template polymer $\mathbf{n}$ made entirely of monomers of type 1. Given the assumed symmetry between the interactions of the two monomer types, and equal concentrations of the monomer baths as used throughout this work, the results apply equally well to any single template sequence. Monomers of type 1 in the copy can simply be interpreted as correct matches and monomers of type 2 as incorrect matches for any template sequence $\mathbf{n}$.

Having defined the model's state space, we now consider state free energies. By analogy with the dimerisation model defined in section 1, we define $\Delta G^{\ominus}_{\text{dim}}$ as the free-energy change of adding any specific monomer to the end of the copy oligomer at standard concentrations, ignoring interactions with any template. The environment contains baths of monomers; a monomer of type s has a constant concentration $[s]$ relative to the standard concentration. The chemical free-energy change for the transition between any

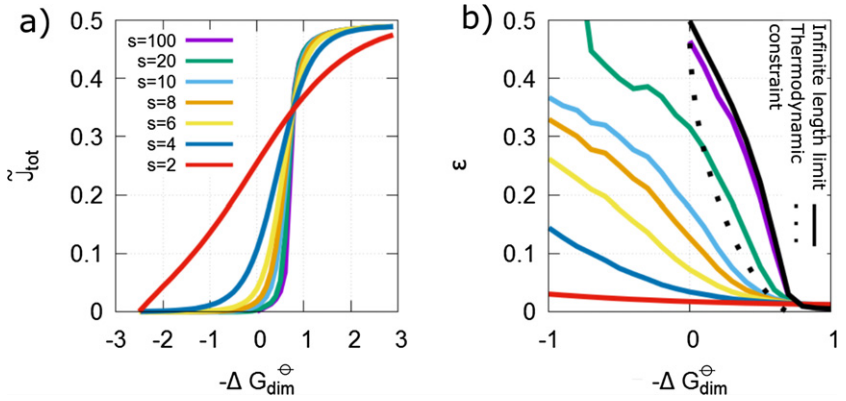

Figure 8. Thermodynamic restrictions in an infinite-length model become kinetic restrictions for longer oligomers. (a) Net flux per unit empty template $\tilde{J}_{\rm tot}$ of oligomer production and (b) the net fraction of error creation for a range of lengths $|\mathbf{s}|$. We vary $\Delta G^{\ominus}_{\rm dim}$, with $[M_{\rm tot}] = 1$, $[S_{\rm tot}]$ chosen to give $\Delta G_{\rm chem} = 0$ at $\Delta G^{\ominus}_{\rm dim} = -5$, $p(s) = t(s)$ unbiased and $\Delta G_{\rm disc} = 8$ at $\Delta G^{\ominus}_{r} + \Delta G^{\ominus}_{w} = 0$. Also plotted in (d) is the thermodynamic constraint on accuracy for an infinite length polymer, set by requiring a non-positive slope of the free-energy profile, $-\epsilon \ln \epsilon - (1-\epsilon)\ln(1-\epsilon) \geqslant \Delta G^{\ominus}_{\rm dim} + \ln 2$, and the actual error rate obtained for an infinite-length copy for these parameters [21]. Short oligomers overcome kinetic barriers to produce copies with $\Delta G^{\ominus}_{\rm dim} > 0$ and ϵ below the infinite-length limit and thermodynamic constraint.

specific sequence $s_1, \ldots, s_l$ and any specific sequence $s_1, \ldots, s_{l+1}$, ignoring any contribution from interactions with the template, is then $\Delta G^{\ominus}_{\rm dim} - \ln[s_{l+1}]$. We then consider the effect of specific interactions with template. Analogously to before, we define $\Delta G^{\ominus}_{r/w}$ as the standard free energy of binding for matched/mismatched monomers and the template at standard concentrations, with $\Delta G_{\rm disc} = \Delta G^{\ominus}_{w} - G^{\ominus}_{r}$ quantifying the additional stability of correct matches. Only the leading monomer is assumed to interact with the template. The overall chemical free-energy change of a given transition is then easily calculable. Each copy extension step from l to $l+1$ is associated with a chemical free-energy change of $\Delta G^{\ominus}_{\rm dim} - \ln[s_{l+1}] + \Delta G_{\rm disc}$ (leading copy monomer goes from $r \to w$); $\Delta G^{\ominus}_{\rm dim} - \ln[s_{l+1}] + 0$ (for $w \to w$ or $r \to r$); or $\Delta G^{\ominus}_{\rm dim} - \ln[s_{l+1}] - \Delta G_{\rm disc}$ (for $w \to r$). These free-energy changes are so simple because each copy extension step does not increase the number of copy monomers interacting with the template (figure 1(c)). Instead, the result is that a different single copy monomer is bound to the template.

A monomer s_1 binding to the first site of the template from solution is associated with a free energy change of $G^{\ominus}_{r/w} - \ln[s_1]$, depending on whether it is a correct match to the first site of the template. Similarly, an oligomer $\mathbf{s}$ binding from solution is associated with a free energy change of $G^{\ominus}_{r/w} - \ln[\mathbf{s}]$, depending on whether it is a correct match to the final site of the template. The differences between these free energies, and those of the polymerisation steps, exemplify the unique role of initiation and termination in setting the overall free-energy change, as illustrated in figure 7.

These free-energy changes constrain the relative propensities of forwards and backwards transitions [29], but a range of kinetic models satify these constraints. In the temporary thermodynamic discrimination model of reference [21], polymerization steps involving the addition of monomer s_i are assumed to occur with the same rate $k[s_i]$, and sequence-based discrimination occurs in the backwards step. We use that parameterization here; in addition, we assume the binding of individual monomers s_i or oligomers $\mathbf{s}$ to an empty template has a rate of $k[s_i]$ or $k[\mathbf{s}]$, with the off-rates fixed by the free-energy change of transition. These assumptions fix all rates, which are given in [23].

To analyse the model we perform stochastic simulations for a range of lengths $|\mathbf{s}|$, varying the free-energy released by backbone formation $\Delta G^{\ominus}_{\rm dim}$ while keeping all other parameters fixed. We use unbiased $m(s)$ and $p(\mathbf{s})$, set $[M_{\rm tot}] = 1$ and choose $[S_{\rm tot}]$ so that $\Delta G_{\rm chem} = 0$ at $\Delta G^{\ominus}_{\rm dim} = 5$; growth is thermodynamically favourable for all sequences when $\Delta G^{\ominus}_{\rm dim} < 5$. However, the slope of the on-template free-energy profile of an unbiased sequence, $\Delta G^{\ominus}_{\rm dim} - \ln[M_{\rm tot}]$, is positive for $\Delta G^{\ominus}_{\rm dim} > 0$. For $5 > \Delta G^{\ominus}_{\rm dim} > 0$, therefore, on-template polymerisation is thus a kinetic barrier to formation of a thermodynamically favourable product.

We calculate the total flux per empty template $\tilde{J}_{\rm tot} = \frac{J_{\rm tot}}{P_{\rm empty}}$ in the steady state and plot it in figure 8. For short oligomers, non-negligible $\tilde{J}_{\rm tot}$ is observed in the region $5 > \Delta G^{\ominus}_{\rm dim} > 0$, indicative of the system surmounting the free energy barriers associated with unfavourable growth on the template. However, as oligomers get longer, the barriers get larger and the kinetics is slowed. Both forward and backwards contributions to $\tilde{J}_{\rm tot}$ are vanishingly small unless $\Delta G^{\ominus}_{\rm dim} < 0$ for $|s| = 100$.

Kinetic barriers not only control the overall production flux per empty template $\tilde{J}_{\mathrm{tot}}$, but also error incorporation. The on-template production of an accurate copy has a more positive slope in its free-energy profile than an unbiased sequence (figure 7(b)). For an infinite-length polymer, this fact provides a thermodynamic constraint on accuracy for $0 > \Delta G^{\ominus}_{\mathrm{dim}} > -\ln 2$ [21]. We plot the fraction of net incorporated monomers that do not match the template, ϵ, in figure 8(b), alongside the thermodynamic constraint on ϵ for infinite-length polymers and the actual error rate obtained for this specific model in the infinite-length limit [21]. Short oligomers can overcome kinetic barriers and beat both the thermodynamic bound and the accuracy obtained in the infinite-length limit; longer oligomers approach the limiting behaviour slowly, with significant differences even at length [20].

The hypothesis that thermodynamic constraints in the infinite-length limit become kinetic barriers for finite-length oligomers, and that these barriers cause kinetic convergence on the behaviour of infinite-length systems for longer oligomers, is therefore confirmed. For shorter oligomers, the edge effects outlined in sections 2 and 3 of this paper have significant effects on the actual distribution of products in a kinetic model, as well as the underlying thermodynamics.

5. Conclusion

In this article we have investigated the copying of finite-length oligomers, with explicit focus on initiation and termination. Accurate copying creates a low entropy sequence, conditioned on the template sequence. In templated self-assembly, this low entropy is costly, but can be compensated for and even favoured in equilibrium due to stabilising interactions of specific copy/template bonds. If these bonds are transient, however, as in the production of a true copy, this compensation cannot play a role in the overall thermodynamics of the process. In an infinite-length model of continuous detachment from behind the leading edge of the copy [21], the low entropy of the product implies a non-equilibrium state. The reduction in the entropy of the copy of $\sim\ln 2$ per monomer relative to a random sequence is not compensated by a discrimination free energy that scales with the length of the copy.

The argument that transient bonds prevent the template from biasing the system towards a low entropy state in equilibrium [21] remains valid once full initiation and termination are considered. However, once the copy has fully detached and mixed with an environment, including oligomers of other sequences, the low entropy of the specific copy oligomer is thermodynamically irrelevant and unexploitable [22]. Since that individual copy can no longer be identified without prior measurement of its sequence, what matters thermodynamically is the relation of that copy sequence the oligomer distribution as a whole, not the entropy of the sequence actually produced. This change in the thermodynamic significance of the copy sequence contributes to the arbitrarily-large offset in the final step of the free energy profile, shown in figure 7.

Thus, as we have shown, the overall thermodynamics of the full copy process does not explicitly depend on accuracy. Instead, the surrounding concentrations of oligomers and monomers set the thermodynamic constraints. Creating outputs that resemble the surrounding oligomers is costly, as is creating outputs unlike the input monomer baths. Arbitrary accuracy can be free-energetically neutral or even actively favourable if the oligomer baths are biased towards other sequences.

However, accuracy does play a role indirectly. Firstly, mixing with other oligomers is the final step, and therefore its thermodynamic consequences are irrelevant whilst a copy is growing on the template. Whilst attached to the template—even if only by the leading site—the copy is distinct from the surrounding oligomer pool and its low entropy is exploitable. For an infinite-length polymer, the associated costs [21] set absolute limits on what is possible. For finite length oligomers, they instead manifest as kinetic barriers; longer oligomers have larger barriers and thus their kinetics converges on the behaviour dictated by the thermodynamic constraints.

Secondly, templates will typically influence their environment. If a template sets its own oligomer environment perfectly, $p(\mathbf{s}) = q(\mathbf{s})$, $\Delta G_{\mathrm{inf}} = D(q||t)$, which reduces to the entropy difference between $t(\mathbf{s})$ and $q(\mathbf{s})$ if $t(\mathbf{s})$ is unbiased. In this case no information is lost upon mixing and accurate copying incurs a cost; the limits derived in reference [21] hold exactly. In general, there is no reason to suppose that $p(\mathbf{s}) = q(\mathbf{s})$. As in a cell, other templates and differential degradation rates may be relevant in setting $p(\mathbf{s})$. Nonetheless, particularly for longer oligomers, sequences common in $q(\mathbf{s})$ are likely to be over-represented in $p(\mathbf{s})$.

If many identical templates are present, then the environmental $p(\mathbf{s})$ will likely be more strongly peaked, and the cost of accuracy higher, than in a system with many distinct templates. Moreover, any template in an environment dominated by the copies of another will experience a relative thermodynamic advantage. This effect would act as a form of 'rubber banding' in evolutionary competition among minimal replicators, and favour virus-like templates invading new environments.

In this work we have derived general results such as equation (2), and analysed explicit models to illustrate these general results. The general results such as equation (2) are valid under relatively weak assumptions, and are independent of the mechanistic details of the copy process. Moreover, although actually calculating the quantities in equation (2) may be impractical for long oligomers, the fundamental insights provided into the thermodynamics remain valid and useful—as demonstrated by our simulations in section 4.1.

The most important assumption is the absence of kinetic proofreading cycles that would consume a variable amount of fuel for each polymer produced and contribute an extra term to the overall free-energy change. Many conclusions in this manuscript would be largely unaffected, however. The existence of kinetic proofreading cycles would effectively provide an unpredictable adjustment to ΔG_{chem} for each oligomer produced; the role of ΔG_{inf} and the copy accuracy in determining overall thermodynamics would be largely unchanged. We note that excluding kinetic proofreading cycles is not equivalent to excluding ancillary fuel molecules with high free energy that drive the process forwards—we simply assume that any such molecules are consumed with the same stoichiometry for each monomer incorporation, and therefore that their free energy can be incorporated into the free energy of polymerisation.

Multiple alternatives exist to the specific models of dimer and oligomer copying used to illustrate these general results. Implicit in both models used here is a mechanism whereby binding together of copy monomers weakens the binding of (at least one of) those monomers to the template, allowing the product to separate and circumvent product inhibition [5–7]. In the case of longer oligomers, the result is detachment of the copy tail from behind its leading edge (figure 1(c)), as is observed in transcription and translation. Since the focus of this paper is thermodynamic constraints, we have not modelled this mechanism in detail, simply assuming that it is possible in a minimal system. This assumption is, however, backed up by recent work [30] in which such a mechanism is demonstrated in a pure nucleic acid-based context. We note that minimal synthetic copying mechanisms such as this may be better suited than highly-evolved biological systems to the exploration of the fundamental limits of copy processes, including the low-cost regime.

Additionally, in the model for the copying of longer templates, we assume that only one copy can be bound to the template at any one time, and that initiation and termination can only occur at the ends of the template. This choice was made to provide the simplest thermodynamically self-consistent model of copying. The mechanistic question of how to engineer reliable initiation and termination, and whether it is advantageous to have only one product on the template at any given time, is the subject of ongoing work.

Finally all non-complementary mismatches are equal and all complementary matches are equal. In practice, this perfect equality will never be obtained—for example, a GC base pair is bound with three hydrogen bonds where a AU base pair is bound with two. Moreover, product sequences do not, in reality, all have the same standard free energy of formation. Exploring the consequences of such differences is the subject of ongoing work. However, we note that the power of transcription and translation is that an essentially arbitrary sequence can be copied, without relying on (and often in spite of) the detailed energetics of the product. Our perfectly symmetric setting is a good starting point to analyse a process that can produce generic products in this way.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Thomas E Ouldridge https://orcid.org/0000-0001-8114-8602

References

[1] Sartori P and Pigolotti S 2015 *Phys. Rev.* X **5** 041039
[2] Alberts B, Bray D, Lewis J, Raff M, Roberts K and Watson J 2002 *Molecular Biology of the Cell* 4th edn (New York: Garland Science)
[3] Murugan A, Zeravcic Z, Brenner M P and Leibler S 2015 *Proc. Natl. Acad. Sci.* **112** 54–9
[4] Johnston I G, Ahnert S E, Doye J P and Louis A A 2011 *Phys. Rev.* E **83** 066105
[5] Vidonne A and Philp D 2009 *Eur. J. Org. Chem.* **2009** 593
[6] Orgel L E 1998 *Trends Biochem. Sci.* **23** 491
[7] Colomb-Delsuc M, Mattia E, Sadownik J W and Otto S 2015 *Nat. Commun.* **6** 7427
[8] Schulman R, Yurke B and Winfree E 2012 *Proc. Natl. Acad. Sci.* **109** 640

[9] Schulman R and Winfree E 2004 Programmable control of nucleation for algorithmic self-assembly *Int. Workshop on DNA-Based Computers* (Berlin: Springer)
[10] Braun D 2004 *Mod. Phys. Lett.* B **18** 775
[11] Zhuo R, Zhou F, He X, Sha R, Seeman N C and Chaikin P M 2019 *Proc. Natl. Acad. Sci.* **116** 1952
[12] Sievers D and Von Kiedrowski G 1994 *Nature* **369** 221
[13] Lincoln T A and Joyce G F 2009 *Science* **323** 1229
[14] Bennett C H 1979 *Biosystems* **11** 85
[15] Cady F and Qian H 2009 *Phys. Biol.* **6** 036011
[16] Andrieux D and Gaspard P 2008 *Proc. Natl. Acad. Sci.* **105** 9516
[17] Sartori P and Pigolotti S 2013 *Phys. Rev. Lett.* **110** 188101
[18] Esposito M, Lindenberg K and Van den Broeck C 2010 *J. Stat. Mech.* **2010** P01008
[19] Ehrenberg M and Blomberg C 1980 *Biophys. J.* **31** 333
[20] Johansson M, Lovmar M and Ehrenberg M 2008 *Curr. Opin. Microbiol.* **11** 141
[21] Poulton J M, ten Wolde P R and Ouldridge T E 2019 *Proc. Natl. Acad. Sci.* **116** 1946
[22] Ouldridge T E and ten Wolde P R 2017 *Phys. Rev. Lett.* **118** 158103
[23] See supplemental material atfor additional data, derivations and discussion.
[24] Hopfield J J 1974 *Proc. Natl. Acad. Sci.* **71** 4135
[25] Gomez M A R and Ibba M 2020 *RNA* **26** 910–36
[26] Deshpande A and Ouldridge T E 2020 *Biol. Cybern.* **114** 653–68
[27] McGrath T, Jones N S, ten Wolde P R and Ouldridge T E 2017 *Phys. Rev. Lett.* **118** 028101
[28] Horowitz J M, Sagawa T and Parrondo J M R 2013 *Phys. Rev. Lett.* **111** 010602
[29] Ouldridge T E 2018 *Nat. Comput.* **17** 3–29
[30] Cabello-Garcia J, Bae W, Stan G-B V and Ouldridge T E 2021 Handhold-mediated strand displacement: a nucleic acid based mechanism for generating far-from-equilibrium assemblies through templated reactions *ACS nano* **15** 3272–83