Length and sequence relaxation of copolymers under recombination reactions

Alex Blokhuis, and David Lacoste

Length and sequence relaxation of copolymers under recombination reactions

Alex Blokhuis and David Lacoste
Gulliver Laboratory, UMR CNRS 7083, PSL Research University, ESPCI, 10 rue Vauquelin,
F-75231 Paris, France

We describe the kinetics and thermodynamics of copolymers undergoing recombination reactions, which are important for prebiotic chemistry. We use two approaches: the first one, based on chemical rate equations and the mass-action law describes the infinite size limit, while the second one, based on the chemical master equation, describes systems of finite size. We compare the predictions of both approaches for the relaxation of thermodynamic quantities towards equilibrium. We find that for some choice of initial conditions, the entropy of the sequence distribution can be lowered at the expense of increasing the entropy of the length distribution. We consider mainly energetically neutral reactions, except for one simple case of non-neutral reactions. *Published by AIP Publishing.*
[http://dx.doi.org/10.1063/1.5001021]

I. INTRODUCTION

Self-assembly plays a central role in biological systems, both for the emergence of life out of non-living matter and for its maintenance. Recent experiments strive to reproduce the sophisticated strategies used by living systems to control self-assembly using complex, typically information-rich elementary bricks.[1] Many new experiments are now possible in this area, thanks to the continuous improvements in experimental techniques of manipulation of nucleic acids and enzymes in micro-fluidic devices. An example of such novel experimental systems are DNA reaction networks allowing for molecular programming and computing.[2,3] Furthermore, with high throughput sequencing techniques, it is now possible to obtain statistical information about the mixtures of nucleic acid sequences with an accuracy and speed out of reach for other types of polymers. Purely artificial copolymers are also being synthetized for applications in information storage.[4,5] The sequence of such polymers can be read or written just like with nucleic acids, albeit with a different chemistry. Clearly, all these experimental techniques are bringing a revolution to biotechnologies.

At the same time, these new experiments also require novel theoretical approaches to account for the rich dynamics displayed by these systems using methods from non-equilibrium Statistical Physics, Thermodynamics, or Information theory. Ideally, one goal would be to build a complete description of the kinetics and thermodynamics of ensembles of polymer sequences undergoing exchange reactions with each other. In view of the complexity of this dynamical system, simplified approaches are needed. Clearly, one important step towards this goal is to understand the dynamics of the length distribution alone disregarding the dynamics of the sequence.

Pioneering theoretical works on reversible polymerization[6,7] were of this type. With the realization in the 1970s that many biopolymers such as actin and microtubules undergo reversible polymerization, new models were built to couple the kinetics of polymerization with the internal energetics of the biopolymer.[8,9] In 2008, a comprehensive model for the thermodynamics of templated copolymerization was developed by Andrieux and Gaspard.[10] This model turned out to be instrumental to understand the general principles of information processing at the molecular scale. While in its original version, only the chemical nature of monomers being added was taken into account, in subsequent work, correlations with the previously added monomers were also included.[11] More recently, the model has also been extended to describe the proof-reading action of exonucleases[12] and sequence heterogeneity effects in the polymerization of DNA or RNA polymerases.[13] In this context, another group also recently investigated the fundamental thermodynamic costs of making polymer copies.[14]

Here, we are not interested in such polymerization reactions, but rather in simpler exchange reactions called recombination reactions. These recombination reactions are reversible and are not necessarily assisted by enzymes such as polymerases. These features make them of interest for prebiotic chemistry, as exchange reactions allow for a large repertoire of sequences to be explored. Inspired by an experimental and theoretical study on the synthesis and degradation of carbohydrates,[15] we have studied in previous work the kinetics and thermodynamics of such reacting polymers.[16] In that work, we considered only one type of monomers, which could either assemble and disassemble by reversible aggregation-fragmentation dynamics or exchange terminal monomer units. The chemical kinetics was described by rate equations following the mass action law, and we assumed a closed system and non-equilibrium initial conditions. Using Stochastic Thermodynamics,[17–20] we have analyzed the conditions under which the mixture dynamically evolves towards an equilibrium state, where detailed balance holds.

In the present paper, we extend that approach by including the sequence of the polymers in the description. We keep otherwise the assumptions of a closed system, non-equilibrium initial conditions, and reversible exchange reactions, all occurring in a well-mixed reactor, in which spatial heterogeneity is neglected. We will consider both energetically neutral and energetically non-neutral reactions.

The outline of this paper is as follows: in Sec. II, we present the two types of exchange reactions, on which we will focus in this paper, which we called chain-exchange and attack-exchange reactions. We then explain briefly the motivations for studying such reactions in the context of prebiotic chemistry. In Sec. III (and, respectively, Sec. IV), we develop a theoretical framework to understand the relaxation of a mixture of polymers undergoing exchange reactions using a deterministic (respectively, stochastic) approach. In Sec. V, we explore the consequences of this approach for the specific case of energetically neutral reactions; while in Sec. VI we study one particular simple case of non-neutral reactions.

II. RECOMBINATION REACTIONS

A. Reaction mechanisms

Chain-exchange and attack-exchange are two examples of recombination reactions, which involve the reversible transfer of a group of subunits between two polymers. Since such reactions conserve the number of chemical bonds between monomers, they are often close to being energetically neutral. The attack-exchange reaction involves the chemical attack of one terminal unit of one chain on a site of the second chain.

Similarly, the chain-exchange reaction involves two polymer chains, which exchange part of their chains.

Exchange reactions can also be thought of as a composition of two reaction steps, such as a fragmentation $\omega_A\omega_B \rightleftharpoons \omega_A + \omega_B$ followed by an addition $\omega_C + \omega_B \rightleftharpoons \omega_C\omega_B$. In the following, it will be advantageous to introduce a specific notation to describe the evolution of sequences according to these reactions (Fig. 1). Monomer sequences are considered to have a distinct polarity (or directionality), as in the case of nucleic acids that have a distinct 5′ and 3′ end. A sequence Ω of length l is composed of $\omega_1\omega_2\ldots\omega_l$. Two subsequent sequences will be noted using a product notation $\omega\omega' = \omega_1\omega_2\ldots\omega_l\omega'_1\omega'_2\ldots\omega'_{l'}$, which is used for the addition of two chains. An inverse sequence is defined as a sequence that is removed, either from the front or from the back, by placing the inverse either in front or on the back of a sequence $\omega\omega'^{-1} = \omega_1\omega_2\ldots\omega_q$. We define a length operator as |.|, which counts the number of elements in a sequence. With

this notation, the attack-exchange may be written as

$$\omega_A\omega_B + \omega_C \rightleftharpoons \omega_C\omega_B + \omega_A. \tag{1}$$
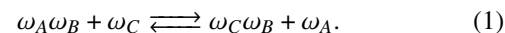
Assuming mass action law, the reaction rates are

$$\begin{aligned} v^{\omega_A\omega_B}_{\omega_C} &= k_{\omega_A\omega_B,\omega_C} N_{\omega_A\omega_B} N_{\omega_C}, \\ v^{\omega_C\omega_B}_{\omega_A} &= k_{\omega_C\omega_B,\omega_A} N_{\omega_C\omega_B} N_{\omega_A}, \end{aligned} \tag{2}$$

where k is the corresponding rate constant, which can be sequence dependent, and N_Ω is the number of polymers of sequence Ω.

Similarly, the chain-exchange reaction drawn in Fig. 2 can be written as

$$\omega_A\omega_B + \omega_C\omega_D \rightleftharpoons \omega_C\omega_B + \omega_A\omega_D, \tag{3}$$
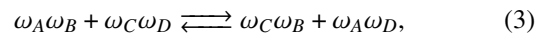
to which we attribute the rates

$$\begin{aligned} v^{\omega_A\omega_B}_{\omega_C\omega_D} &= k_{\omega_A\omega_B,\omega_C\omega_D} N_{\omega_A\omega_B} N_{\omega_C\omega_D}, \\ v^{\omega_A\omega_D}_{\omega_C\omega_B} &= k_{\omega_A\omega_D,\omega_C\omega_B} N_{\omega_A\omega_D} N_{\omega_C\omega_B}. \end{aligned} \tag{4}$$

When the forward and backward rate constants $k_{\omega_A\omega_B,\omega_C\omega_D}$ and $k_{\omega_A\omega_D,\omega_C\omega_B}$ are equal, there is no change of standard free energy, which implies a compensation between standard enthalpy and entropy as detailed in Subsection III B.

An important constraint for both reactions of Eqs. (1) and (3) is that we exclude the formation of any species of zero length. This means that the total number of chains $N = \sum_\Omega N_\Omega$ is a conserved quantity for both dynamics. In other words, there is a minimum length of chains $l_{min} = 2$ for chain-exchange reactions while $l_{min} = 1$ for attack-exchange reactions. In addition, in both exchange reactions, the first monomer is never displaced, which leads to a conservation law for the composition of the first monomer. For chain-exchange, such a law also exists for terminal monomers because they always remain in a terminal position.

B. Prebiotic context

In studies on prebiotic chemistry, recombination reactions are being more and more considered potential key players before the emergence of truly self-replicating systems.[21] Indeed, recombination reactions do not require complex enzymes,[22,23] nor do they require an energy source or abundant monomer supplies. At the same time, their dynamics is sufficiently rich that it can allow for a broadening of the length distribution of the polymers and the apparition of a primitive form of inheritability of their sequence,[23] which was until recently believed to be only possible in systems evolving by template-assisted polymerization.[24]

All these features make recombination reactions promising candidates in prebiotic scenarios, as a means to explore the
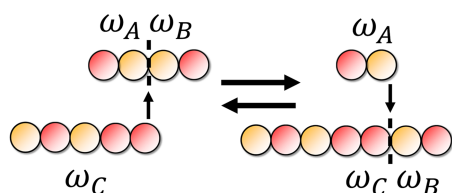

FIG. 1. Representation of attack-exchange reaction: $\omega_A\omega_B + \omega_C \rightleftharpoons \omega_C\omega_B + \omega_A$ for the case that two monomer types are present: $m = 2$.
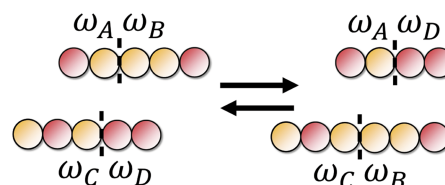

FIG. 2. Representation of chain-exchange reaction: $\omega_A\omega_B + \omega_C\omega_D \rightleftharpoons \omega_A\omega_D + \omega_C\omega_B$.

functional space of polymer sequences given prebiotic conditions (no cellular machinery, low specificity/control, unreliable source of energy, or monomers). In RNA-world scenarios, this exploration could eventually lead to the emergence of catalytically active RNA species, so-called "ribozymes." As pointed out by Lehman, exchange reactions function very much like sexual recombination in chromosomes, which accounts for most of today's natural variation. This is in contrast with template-directed polymerizations that resemble, in this respect, asexual cloning.[25]

In the context of nucleic acid chemistry, several reactions could qualify for recombination reactions: An attack-exchange reaction occurs when the terminal hydroxyl group (possibly modified) of one nucleic acid attacks a phosphodiester group of another nucleic acid polymer and is typically aided by sequence complementarity.[22,23] This nucleophilic attack is a simple transesterification, in which the number of phosphodiester bonds is conserved. In modern biology, this reaction is an important part of mRNA splicing pathways, in which a pre-mRNA containing intron and exon sequences is finalized by the removal of introns. Key steps for such a process are (I) cleavage at a 5′ splicing site, followed by (II) transesterification at a 3′ splicing site, in which an exon displaces an intron, yielding a final RNA consisting of solely exon sequences. These reactions are controlled by a number of ribozymes and enzymes, and a single pre-mRNA often presents a whole repertoire of alternative splicing pathways.[26] Protein splicing is a similar process, which often proceeds through a similar transesterification step.[27] A synthetic example of a ribozyme performing exchange reactions is the Azoarcus ribozyme.[28]

In synthetic chemistry, an attack-exchange is mediated by any reaction with a "trans-" prefix (e.g., transesterification, transamination, and transamidation). The type of reaction we call chain-exchange is referred to as metathesis (e.g., olefin metathesis, disulfide metathesis). The interest of sequence exploration is not limited to RNA. For example, a recent example shows how certain tripeptide sequences can lead to an assembly of functional ferrodoxin clusters.[29] In the last few decades, the chain-exchange reaction has become essential in synthetic chemistry, which culminated in the 2005 Nobel Prize in Chemistry for metathesis methods. In particular, the metathesis of olefins has become an invaluable tool for the chemist.[30]

In 2012, a novel class of polymers called vitrimers were discovered by Leibler's group, which capitalize on the dynamic properties provided by exchange reactions.[31,32] By now, vitrimers have been developed employing various exchange reactions, such as disulfide metathesis, transamination,[33] transalkylation, and many others, for which we refer to a recent mini review in Ref. 34.

III. EXCHANGE-REACTION THERMODYNAMICS

A. Equilibrium thermodynamics

We now present a thermodynamic framework to describe the dynamics of a polymer mixture undergoing chain-exchange reactions. There will be two steps, here we discuss equilibrium thermodynamics features then in Sec. III B, we

will present the non-equilibrium thermodynamic ones. Since the calculations for the attack-exchange reaction would be rather similar, we will simply point out differences between the two dynamics when appropriate. We assume that the mixture contains m different monomer types $\{0, 1, 2 \ldots, m-1\}$, with $m > 1$ so that polymer sequences can be defined.

In our modeling of the chemistry, we do not include the solvent explicitly in the description. We refer the reader to Ref. 16 for an illustration of an explicit inclusion of the solvent in the kinetics and thermodynamics of polymerization models. We recall that the two exchange reactions we are interested in conserve the following quantity $N = \sum_\Omega N_\Omega$, which represents the total concentration of chains (including monomers). Therefore, we define the polymer fraction of sequence Ω as

$$y_\Omega = \frac{N_\Omega}{N}, \tag{5}$$

which obeys the normalization condition $\sum_\Omega y_\Omega = 1$. We assume that the solution is dilute and thus the chemical potentials of all present species follow the form

$$\mu_\Omega = \mu_\Omega^\circ + k_B T \ln y_\Omega, \tag{6}$$

where T is the temperature. The enthalpy of the solution can be expressed in terms of h_Ω° the standard enthalpy of a sequence Ω as

$$H = \sum_\Omega N_\Omega \, h_\Omega^\circ. \tag{7}$$

Likewise, the entropy can be defined in this manner,

$$S = \sum_\Omega N_\Omega (s_\Omega^\circ - k_B \ln y_\Omega), \tag{8}$$

where s_Ω° represents the internal contribution of the entropy associated with other degrees of freedom different from Ω and not described here. We will also use the system entropy per chain $\mathcal{S}$ defined as

$$\mathcal{S} = \frac{S}{N} = \sum_\Omega y_\Omega (s_\Omega^\circ - k_B \ln y_\Omega). \tag{9}$$

Let us define $G = H - TS$ as the Gibbs free energy. Using $\mu_\Omega = h_\Omega - T s_\Omega$, we find $G = \sum_\Omega N_\Omega \mu_\Omega = \sum_\Omega N_\Omega (\mu_\Omega^\circ - k_B T \ln y_\Omega)$. In the remainder of this paper, we will take $k_B = 1$.

B. Non-equilibrium thermodynamics

We now move to a description of the non-equilibrium thermodynamic part of the problem, and to do that we introduce the kinetic rate equation for the concentration of chains with sequence Ω is

$$\dot{N}_\Omega = \sum_{\omega_A = \Omega \omega_B^{-1}} \sum_{\omega_C} \sum_{\omega_D} \left[v_{\omega_C \omega_B}^{\omega_A \omega_D} - v_{\omega_C \omega_D}^{\omega_A \omega_B} \right]. \tag{10}$$

The kinetic constant is taken to be dependent on the exact sequences and on the sites of splitting. The chain-exchange reaction exchanges chemical bonds between subsequences of nonzero length. As such, the set of subsequences we consider cannot be empty ($\omega \neq \emptyset$) and a total sequence is at least of length 2. For convenience, we choose to make this instruction implicit.

The second term is equivalent to the back reaction of the first term. When summing over all possible sequences Ω, the

first sequence sum turns into a sum over subsequences ω_A and ω_B [all distinct ordered pairs (ω_A, ω_B) are generated],

$$\sum_{\Omega} \sum_{\omega_A = \Omega \omega_B^{-1}} = \sum_{\omega_A} \sum_{\omega_B}, \tag{11}$$

which generates the symmetry $\sum_{\Omega} \dot{N}_{\Omega} = -\sum_{\Omega} \dot{N}_{\Omega}$. This of course implies the conservation of the number of chains $\sum_{\Omega} \dot{N}_{\Omega} = 0$. The entropy production rate Σ of an ensemble of chemical reactions, assumed to be elementary (there should be no hidden chemical reactions), taking the form[35]

$$\Sigma = \sum_{k} (v_k^+ - v_k^-) \ln \left(\frac{v_k^+}{v_k^-} \right) \geq 0, \tag{12}$$

where v_k^+, v_k^- are, respectively, forward and backward reaction rates of the kth reaction. In the specific case of chain-exchange reactions, this becomes

$$\Sigma = \frac{1}{4} \sum_{\Lambda} \left[v_{\omega_C \omega_B}^{\omega_A \omega_D} - v_{\omega_C \omega_D}^{\omega_A \omega_B} \right] \ln \left(\frac{v_{\omega_C \omega_D}^{\omega_A \omega_B}}{v_{\omega_C \omega_B}^{\omega_A \omega_D}} \right), \tag{13}$$

where the sum is carried out over Λ, which represents an arbitrary set of four sequences of the form $\{\omega_A, \omega_B, \omega_C, \omega_D\}$. The factor 4 can be understood as the cardinal of a discrete group $\mathcal{G}$ acting on elements of Λ. This group contains the following 4 elements: $\mathcal{G} = \{I, \chi, \pi, \rho\}$, where I is the identity, χ presents the exchange $\omega_A \rightarrow \omega_C$, π presents the exchange $\omega_B \rightarrow \omega_D$, and ρ presents the combined exchange $\omega_A \rightarrow \omega_C$ and $\omega_B \rightarrow \omega_D$. Similarly, for attack-exchange, the relevant group $\mathcal{H}$ contains instead the elements: $\mathcal{H} = \{I, \chi\}$. Since the cardinal of $\mathcal{H}$ is 2 instead of 4 for $\mathcal{G}$, the equivalent of Eq. (13) for attack-exchange should contain a factor 2 in the place of the factor 4.

Detailed balance should hold at equilibrium, which provides the following relation:

$$k_{\omega_A \omega_B, \omega_C \omega_D} y_{\omega_A \omega_B}^{eq} y_{\omega_C \omega_D}^{eq}$$
$$= k_{\omega_A \omega_D, \omega_C \omega_B} y_{\omega_A \omega_D}^{eq} y_{\omega_C \omega_B}^{eq}. \tag{14}$$

Then, the condition $\Delta \mu = 0$ together with Eq. (6) leads to

$$T \ln \left(\frac{y_{\omega_A \omega_B}^{eq} y_{\omega_C \omega_D}^{eq}}{y_{\omega_C \omega_B}^{eq} y_{\omega_A \omega_D}^{eq}} \right) = -\Delta \mu_{\omega_A \omega_B, \omega_C \omega_D}^{\circ}$$
$$= \mu_{\omega_A \omega_D}^{\circ} + \mu_{\omega_C \omega_B}^{\circ} - \mu_{\omega_C \omega_D}^{\circ} - \mu_{\omega_A \omega_B}^{\circ}. \tag{15}$$

Combining this relation with detailed balance (14), one obtains

$$T \ln \left(\frac{k_{\omega_C \omega_D, \omega_A \omega_B}}{k_{\omega_A \omega_D, \omega_C \omega_B}} \right) = -\Delta \mu_{\omega_A \omega_B, \omega_C \omega_D}^{\circ}. \tag{16}$$

We emphasize that for energetically neutral reactions, the forward and backward rates are equal, which implies $\Delta \mu^{\circ} = 0$. In that case, a compensation between standard entropy and enthalpy must occur, since $\Delta h^{\circ} = T \Delta s^{\circ}$.

If we now calculate the time evolution of the enthalpy H, we obtain

$$\frac{dH}{dt} = \sum_{\Lambda} \left[v_{\omega_C \omega_B}^{\omega_A \omega_D} - v_{\omega_C \omega_D}^{\omega_A \omega_B} \right] h_{\omega_A \omega_B}^{\circ}$$
$$= \frac{1}{4} \sum_{\Lambda} \left[v_{\omega_C \omega_B}^{\omega_A \omega_D} - v_{\omega_C \omega_D}^{\omega_A \omega_B} \right] \Delta h_{\omega_A \omega_B, \omega_C \omega_D}^{\circ}, \tag{17}$$

where we used the symmetry to write the evolution in single-reaction enthalpy changes

$$\Delta h_{\omega_A \omega_B, \omega_C \omega_D}^{\circ} = h_{\omega_A \omega_B}^{\circ} + h_{\omega_C \omega_D}^{\circ}$$
$$- h_{\omega_A \omega_D}^{\circ} - h_{\omega_C \omega_B}^{\circ}. \tag{18}$$

Similarly, for the entropy, we obtain

$$\frac{dS}{dt} = \frac{1}{4} \sum_{\Lambda} \left[v_{\omega_C \omega_B}^{\omega_A \omega_D} - v_{\omega_C \omega_D}^{\omega_A \omega_B} \right]$$
$$\times \left[\Delta s_{\omega_A \omega_B, \omega_C \omega_D}^{\circ} - \ln \left(\frac{y_{\omega_A \omega_B} y_{\omega_C \omega_D}}{y_{\omega_A \omega_D} y_{\omega_C \omega_B}} \right) \right]. \tag{19}$$

We can combine Eqs. (15), (17), and (19) to get

$$\frac{dG}{dt} = \frac{T}{4} \sum_{\Lambda} \left[v_{\omega_C \omega_D}^{\omega_A \omega_B} - v_{\omega_C \omega_B}^{\omega_A \omega_D} \right]$$
$$\times \ln \left(\frac{y_{\omega_C \omega_B} y_{\omega_A \omega_D} y_{\omega_A \omega_B}^{eq} y_{\omega_C \omega_D}^{eq}}{y_{\omega_C \omega_B}^{eq} y_{\omega_A \omega_D}^{eq} y_{\omega_A \omega_B} y_{\omega_C \omega_D}} \right). \tag{20}$$

Using detailed balance (14) into Eq. (20), one recovers the previous expression defined in Eq. (13) for the entropy production rate Σ,

$$-\frac{1}{T} \frac{dG}{dt} = \Sigma = -\sum_{\Omega} \dot{N}_{\Omega} \ln \left(\frac{N_{\Omega}}{N_{\Omega}^{eq}} \right) \geq 0. \tag{21}$$

Since $G = H - TS$, this equation is equivalent to $\dot{S} = \Sigma + \dot{H}/T$, which expresses the second law of thermodynamics for a closed system. As expected, the heat released by the system into the environment Q is the change of enthalpy $Q = \Delta H$. Equation (21) is important to guarantee that the chemical system reaches a unique equilibrium state on long times.[16]

C. Decomposition of the entropy production

Here, we split the entropy production of the polymer mixture into two contributions, where the first one represents the contribution of the various polymer lengths, while the second one represents that of their sequences. Using Eqs. (5) and (21), we can rewrite the entropy production rate Σ in terms of polymer fractions,

$$\Sigma = -N \frac{d}{dt} \sum_{\Omega} y_{\Omega} \ln \left(\frac{y_{\Omega}}{y_{\Omega}^{eq}} \right)$$
$$= -N \frac{d}{dt} \sum_{\Omega} y_{\Omega} \left(\frac{\mu_{\Omega}^{\circ}}{T} + \ln y_{\Omega} \right). \tag{22}$$

Since the polymer fractions y_{Ω} for all sequences Ω are normalized, y_{Ω} can be interpreted as the probability to observe a chain of sequence Ω among all possible sequences. Furthermore, since the polymer of sequence Ω has only one possible length, namely, $l = |\Omega|$, that probability to observe a polymer with sequence Ω can be denoted equivalently $P_{\Omega, l}(t)$ because the length is a redundant variable. At any time t, we have therefore the identification

$$y_{\Omega}(t) = P_{\Omega, l}(t). \tag{23}$$

To proceed, we then factorize $P_{\Omega, l}(t)$ in the following way:

$$P_{\Omega, l}(t) = Y_l(t) \, U_{l, \Omega}(t), \tag{24}$$

with $Y_l(t)$ the probability distribution of polymer length at time t, and $U_{l,\Omega}(t)$ the conditional probability distribution of the sequence, and conditional on the length l. The distributions Y_l and $U_{l,\Omega}$ are normalized: $\sum_l Y_l(t) = 1$, and $\sum_\Omega U_{l,\Omega}(t) = 1$ provided the sum is restricted to all chains that have a length l.

The inspiration for the factorization in Eq. (24) comes from the work of Andrieux and Gaspard,[10] where a similar relation has been used to model the thermodynamics of copolymerization of a single polymer. Let us emphasize, however, important differences between our work and this reference. In the work of Andrieux and Gaspard, a single polymer grows and shrinks by the addition or removal of single units at one of its end, which leads to a steady growth regime on long times. In that steady growth regime, the polymer has a time-dependent length distribution $Y_l(t)$ but a stationary sequence distribution for length l, $U_{l,\Omega}$. In contrast, we do not have a steady growth regime here, we consider a polymer mixture rather than a single polymer. Further, our polymers do not grow or shrink only by the ends but undergo exchange reactions, which eventually make the system relax to equilibrium instead of reaching a non-equilibrium steady state as in the work of Andrieux and Gaspard.

Unless indicated otherwise, the distributions Y_l and $U_{l,\Omega}$ are assumed to be time-dependent. For attack-exchange, however, the sequence relaxes more slowly than the length (as shown in Appendix A). Therefore, there is a specific time window in which all the time dependence is carried by $U_{l,\Omega}$ and not Y_l: $P_{\Omega,l}(t) = Y_l U_{l,\Omega}(t)$.

Let us now go back to the general case. Using Eqs. (22)–(24), we deduce a splitting of the entropy production rate into three contributions,

$$\Sigma = -N \frac{d}{dt} \left[\sum_l Y_l \ln Y_l + \sum_{\Omega,l} Y_l U_{l,\Omega} \ln U_{l,\Omega} \right.$$
$$\left. + \sum_{\Omega,l} Y_l U_{l,\Omega} \frac{\mu_\Omega^\circ}{T} \right]. \tag{25}$$

The various terms in this decomposition are:

- The first term: $\sum_l Y_l \ln Y_l$ represents the disorder in the length distribution Y_l (or length entropy).
- The second term: $\sum_{\Omega,l} Y_l U_{l,\Omega} \ln U_{l,\Omega}$ represents the disorder in the distribution of sequences (or sequence entropy). Importantly, this term is weighted by the length distribution Y_l and therefore introduces a coupling between length and sequence distributions. As a result, one expects that the dominant contribution to this sequence entropy will come from short sequences.
- The final contribution: $\sum_{\Omega,l} Y_l U_{l,\Omega} \, \mu_\Omega^\circ/T$ comes from the standard free energy change of each species. If we choose μ_Ω° such that our reactions are energetically neutral: $\Delta\mu^\circ = \mu_{\omega_A\omega_B}^\circ + \mu_{\omega_C\omega_D}^\circ - \mu_{\omega_C\omega_B}^\circ - \mu_{\omega_A\omega_D}^\circ = 0$, this term vanishes. This term can be split further into two using $\mu^\circ = h^\circ - Ts^\circ$. Two terms will appear, $\sum_{\Omega,l} Y_l U_{l,\Omega} \, h_\Omega^\circ$, which corresponds to the heat exchanged with the surrounding medium and $\sum_{\Omega,l} Y_l U_{l,\Omega} \, s_\Omega^\circ$ which corresponds to an internal entropy contribution to Σ.

Given an initial distribution $Y_l^I, U_{l,\Omega}^I$ and final distribution $Y_l^F, U_{l,\Omega}^F$, the total entropy production per chain $\Delta\mathcal{S}_{tot}$ in that transformation follows from (25),

$$\Delta\mathcal{S}_{tot} = \sum_l \left(Y_l^I \ln Y_l^I - Y_l^F \ln Y_l^F \right)$$
$$+ \sum_{\Omega,l} \left(Y_l^I U_{l,\Omega}^I \ln U_{l,\Omega}^I - Y_l^F U_{l,\Omega}^F \ln U_{l,\Omega}^F \right)$$
$$+ \sum_{\Omega,l} \left(Y_l^I U_{l,\Omega}^I - Y_l^F U_{l,\Omega}^F \right) \frac{\mu_\Omega^\circ}{T}. \tag{26}$$

To derive this result, we have used mainly the detailed balance condition and the two conservation laws introduced earlier for the total number of chains and of monomers.

IV. STOCHASTIC THERMODYNAMICS FRAMEWORK

Section III relied on mass action laws and kinetic rate equations, which are appropriate in the thermodynamic limit when the number of chains $N \to \infty$. In a small system where fluctuations matter, a different approach is needed based on Stochastic Thermodynamics.[17–19] We define a state $\mathbf{n} = \{n_{\Omega_1}, n_{\Omega_2}, n_{\Omega_3}, \ldots\}$, as a vector containing the numbers of each polymer (distinguished by their sequence and length) present in the system. The probability to be in a given state $\mathbf{n}$, $P(\mathbf{n})$, which obeys the following master equation:[36]

$$\frac{dP(\mathbf{n})}{dt} = \sum_{\mathbf{n}'} [W_{\mathbf{n}' \to \mathbf{n}} P(\mathbf{n}') - W_{\mathbf{n} \to \mathbf{n}'} P(\mathbf{n})], \tag{27}$$

where $W_{\mathbf{n} \to \mathbf{n}'}$ is the transition rate to jump from $\mathbf{n}$ to $\mathbf{n}'$. Given the size of the sequence space, this equation is difficult to solve analytically, but we can nevertheless derive some useful results from it.

It is important to appreciate that the states $\mathbf{n}$ have an internal degeneracy $z(\mathbf{n})$, which follows from all the allowed permutations among polymer sequences compatible with that state,

$$z(\mathbf{n}) = \frac{N!}{n_{\Omega_1}! n_{\Omega_2}! \ldots n_{\Omega_n}! \ldots} = \frac{N!}{\prod_\Omega (n_\Omega!)}. \tag{28}$$

The analogues of the ensemble averaged number of polymers of sequence Ω, N_Ω and of the entropy S introduced in Sec. III A are the stochastic particle number n_Ω and the stochastic entropy s. The connection between the two descriptions is that

$$N_\Omega = \langle n_\Omega \rangle, \tag{29}$$
$$S = \langle s \rangle, \tag{30}$$

where the average is taken with respect to the distribution $P(\mathbf{n})$. Now, the expression of the stochastic entropy s is[37]

$$s(\mathbf{n}) = -\ln P(\mathbf{n}) + \ln z(\mathbf{n}) + s^\circ(\mathbf{n}), \tag{31}$$

where the first term on the right-hand side gives after averaging over the distribution of $\mathbf{n}$ the Shannon entropy of that distribution, the second term is the contribution of the degeneracy while the last term is the internal entropy coming from non-described molecular degrees of freedom. The precise definition of that last term is

$$s°(\mathbf{n}) = \sum_\Omega n_\Omega s_\Omega°, \qquad (32)$$

in terms of $s_\Omega°$, the intensive standard entropy of formation introduced in Eq. (9).

Assuming the reaction $\mathbf{n} \to \mathbf{n}'$ is elementary (i.e., the two vectors differ by only one recombination reaction among two of their components), the detailed balance condition is

$$\frac{W_{\mathbf{n}\to\mathbf{n}'}}{W_{\mathbf{n}'\to\mathbf{n}}} = \frac{z(\mathbf{n}')}{z(\mathbf{n})} \exp(-\beta\Delta\mu°), \qquad (33)$$

where $\beta = 1/T$ and $\Delta\mu°$ is the chemical potential difference of the elementary exchange reaction introduced in Sec. III B. We recall that the latter may be split into $\Delta\mu° = \Delta h° - T\Delta s°$.

In the absence of degeneracy, the ratio $\ln W_{\mathbf{n}\to\mathbf{n}'}/W_{\mathbf{n}'\to\mathbf{n}}$ would correspond to the stochastic heat transferred from the system to the reservoir during that transition. However, in present case, due to the degeneracy, the correct definition of the stochastic heat, δq is

$$-\beta\delta q = \ln \frac{W_{\mathbf{n}\to\mathbf{n}'}}{W_{\mathbf{n}'\to\mathbf{n}}} - \ln \frac{z(\mathbf{n}')}{z(\mathbf{n})} - \Delta s°. \qquad (34)$$

Using (33) and (34), it follows immediately that

$$\delta q = \Delta h°. \qquad (35)$$

When summing (35) over all transitions, we obtain the total heat $q(t)$ exchanged with the heat bath, at time t, in the form of a sum over all past events indexed by j,

$$q(t) = \sum_j \delta q_j. \qquad (36)$$

According to the second law of Stochastic Thermodynamics,[17,19] the total entropy production on this trajectory is

$$\Delta s_{tot} = \Delta s + \Delta s_m, \qquad (37)$$

where Δs is the change of system entropy between the final and initial states and Δs_m the change in medium entropy. The latter is fundamentally associated to the heat defined above by $\Delta s_m = -\beta q$.

Given Eq. (31), the difference of system entropy is

$$\Delta s = \ln \frac{P(\mathbf{n}^I)}{P(\mathbf{n}^F)} + \ln \frac{z(\mathbf{n}^F)}{z(\mathbf{n}^I)} + s°(\mathbf{n}^F) - s°(\mathbf{n}^I), \qquad (38)$$

which when combined with Eqs. (32)–(34), leads to the expected central result that the total entropy production is the ratio of the probability of forward paths to that of backward paths,

$$\Delta s_{tot} = \ln \frac{P(\mathbf{n}^I)W_{\mathbf{n}^I\to\mathbf{n}^1}\dots W_{\mathbf{n}^{F-1}\to\mathbf{n}^F}}{P(\mathbf{n}^F)W_{\mathbf{n}^1\to\mathbf{n}^I}\dots W_{\mathbf{n}^F\to\mathbf{n}^{F-1}}}. \qquad (39)$$

The contribution due to degeneracy can be further split as

$$\frac{1}{N} \ln \frac{z(\mathbf{n}^F)}{z(\mathbf{n}^I)} = \frac{1}{N} \ln \frac{\prod_\Omega n_\Omega^I!}{\prod_\Omega n_\Omega^F!} = \Delta s_L + \Delta s_\omega, \qquad (40)$$

with Δs_L being the length entropy per chain and Δs_ω being the weighted sequence entropy per chain of a finite system,

$$\Delta s_L = \frac{1}{N} \ln \frac{\prod_l n_l^I!}{\prod_l n_l^F!},$$
$$\Delta s_\omega = \frac{1}{N} \ln \frac{\prod_\Omega n_\Omega^I!}{\prod_\Omega n_\Omega^F!} - \frac{1}{N} \ln \frac{\prod_l n_l^I!}{\prod_l n_l^F!}. \qquad (41)$$

A. Connection to the macroscopic approach

It is interesting to check that the above framework is compatible with the expressions obtained previously in the macroscopic approach. We assume that there is no distribution of the initial condition. Therefore, in the change of stochastic system entropy defined in Eq. (38), we need to focus on $P(\mathbf{n}^F)$ since $P(\mathbf{n}^I) = 1$ and therefore $\ln P(\mathbf{n}^I) = 0$. In order to evaluate $P(\mathbf{n}^F)$, let us assume that the system has reached equilibrium at the final time. For a macroscopic system that probability distribution takes the equilibrium form

$$P(\mathbf{n}^F) = z(\mathbf{n}^F)\prod_\Omega (y_\Omega)^{n_\Omega^F}, \qquad (42)$$

where we have used the definition of the degeneracy factor in Eq. (28) and the conservation law of the number of chains $\sum_\Omega n_\Omega = N$. To make the connection with the macroscopic description, we can show that the polymer fractions y_Ω previously defined in Eq. (5) must also be the ensemble average of n_Ω divided by N,

$$y_\Omega = \frac{\langle n_\Omega^F \rangle}{N}, \qquad (43)$$

where the average is taken with respect to the equilibrium distribution of Eq. (42). Now, by reporting Eq. (42) into Eq. (31), one finds

$$s(\mathbf{n}^F) = -\sum_\Omega n_\Omega^F \ln y_\Omega + s°(\mathbf{n}^F). \qquad (44)$$

When this expression is averaged over the equilibrium distribution of Eq. (42), one recovers using Eqs. (29) and (32) the familiar expression of the entropy introduced in the equilibrium thermodynamics section, namely, Eq. (8).

Let us discuss the connection to the macroscopic approach for the separate contributions of length and sequence. We start by using Stirling's approximation in Eq. (41), $\ln n!$ $= n \ln n - n + O(\ln n)$. In this limit, one recovers the expected contributions to the entropy:

$$\Delta s_L \approx \sum_l \left[\frac{n_l^I}{N} \ln \frac{n_l^I}{N} - \frac{n_l^F}{N} \ln \frac{n_l^F}{N}\right],$$
$$\Delta s_\omega \approx \sum_{l,\Omega} \left[\frac{n_\Omega^I}{N} \ln \frac{n_\Omega^I}{N} - \frac{n_\Omega^F}{N} \ln \frac{n_\Omega^F}{N}\right] - \Delta s_L. \qquad (45)$$

In the thermodynamic limit, the probability distribution of n_Ω becomes peaked around the value $\langle n_\Omega \rangle = N_\Omega$. By replacing n_Ω by N_Ω and n_l by N_l and using the definitions: $N_\Omega = NY_l U_{l,\Omega}$ and $N_l = NY_l$, in Eq. (45), one recovers precisely the first two terms in (26). In this limit, the n_Ω becomes deterministic, therefore, the first term in Eq. (38) becomes negligible.

Finally, we note that the heat per polymer is:

$$\frac{q}{N} = \sum_{l,\Omega} \left[Y_l^F U_{l,\Omega}^F - Y_l^I U_{l,\Omega}^I\right] h_\Omega°, \qquad (46)$$

while the internal entropy part is similarly

$$\mathcal{S}° = \sum_{l,\Omega} \left[Y_l^F U_{l,\Omega}^F - Y_l^I U_{l,\Omega}^I\right] s_\Omega°. \qquad (47)$$

By combining Eqs. (45)–(47), we see that we recover all the terms in the entropy production of Eq. (26) obtained in the macroscopic approach.

V. SIMULATIONS WITH ENERGETICALLY NEUTRAL REACTIONS

In a mean-field description, a mixture of well stirred reacting polymers undergoing exchange reactions is simulated with a Gillespie (Dynamic Monte Carlo) algorithm.[38] In this section, we study numerically the relaxation of thermodynamic quantities Δs_L and Δs_ω for such a system. The simulation uses a list of length N, in which each entry corresponds to a sequence stored as a string. This list is updated for every subsequent reaction step and changes in Δs_L and Δs_ω are calculated from Eq. (41).

For energetically neutral chain-exchange reactions, the forward and backward rates are equal: $k_{\omega_A\omega_B,\omega_C\omega_D} = k_{\omega_A\omega_D,\omega_C\omega_B}$. For simplicity, we choose these reaction rates to be constant independent of the sequence: $k_{\omega_A\omega_B,\omega_C\omega_D} = 1$.

A. Equilibrium length distributions

We will first study the length distribution $Y_l = N_l/N$, with N_l the number of polymers of length l, and N the total number of polymers. We have two separate conservation laws for the number of chains: $\sum_{l=l_{min}}^{\infty} N_l = N$ and for the number of monomers (mass conservation): $\sum_{l=l_{min}}^{\infty} lN_l = M$, with l_{min} the length of the shortest possible species. Now, detailed balance imposes $N_{l_A}N_{l_B} = N_{l_C}N_{l_D}$ with $l_A + l_B = l_C + l_D$, which leads to an exponential length distribution: $N_l = A(B)^{l-l_{min}}$, where A and B are constants depending on the mechanism. Solving the algebraic equations for N_l for chain-exchange where $l_{min} = 2$ yields

$$N = \frac{A}{1-B}, \quad M = -\frac{AB^2(B-2)}{B^2(1-B)^2} = \frac{A(B-2)}{(1-B)^2}, \quad (48)$$

from which we find

$$A = \left(\frac{N}{\frac{M}{N}-1}\right), \quad B = \left(\frac{\frac{M}{N}-2}{\frac{M}{N}-1}\right). \quad (49)$$

We thus have an expression for Y_l^{eq},

$$Y_l^{eq} = \frac{1}{\frac{M}{N}-1}\left(\frac{\frac{M}{N}-2}{\frac{M}{N}-1}\right)^{l-2}. \quad (50)$$

For attack-exchange, $l_{min} = 1$ and a similar calculation leads to

$$Y_l^{eq} = \frac{N}{M}\left(1-\frac{N}{M}\right)^{l-1}. \quad (51)$$

Such exponential length distributions were already obtained long ago by Flory,[6] Blatz, and Tobolsky[7] in their pioneering work on reversible polymerization.

It is important to appreciate that these equilibrium distributions also hold when the polymers contain different types of monomers (i.e., when $m \neq 1$). Indeed, the conservation laws and detailed balance conditions hold and fix the equilibrium length distribution independently of the chemical composition; therefore, they cannot depend on m. This may no longer be the case, however, when there is an energy function attached to the polymers depending specifically on the chemical nature of the monomers.

B. Equilibrium sequence distributions

Let us now discuss the equilibrium state of sequences. When there is no energy function and when monomers are equally abundant, all m^l possible sequences of length l are equiprobable, thus: $U_{l,\Omega}^{eq} = 1/m^l$. At equilibrium, the weighted sequence disorder for both mechanisms reaches the same maximum value,

$$-\sum_{\Omega,l} Y_l^{eq} U_{l,\Omega}^{eq} \ln U_{l,\Omega}^{eq} = -\sum_l Y_l^{eq} \ln\left(\frac{1}{m^l}\right)$$

$$= -\sum_l lY_l^{eq} \ln\left(\frac{1}{m}\right) = \frac{M}{N}\ln(m). \quad (52)$$

C. Kinetics

In this system, we can consider the following relaxation times as shown in Table I: (i) the mean reaction time is $1/k$, (ii) the waiting time τ_r is the time it takes to perform the next chemical reaction. For instance for attack-exchange, this time is the mean reaction time divided by the total number of reactions. Since each reaction involves one terminal unit of one polymer and another polymer from the pool, the number of reactions equals the number of bonds, $M - N$, times the number of polymers N. Then, (iii) is the relaxation time of the length τ_l, which is defined as follows. From the kinetic rate equations, it can be shown that the number of polymers of length l, N_l can be written as a sum of exponentials, and τ_l is the longest relaxation time in that decomposition. Then, (iv) a characteristic time for sequence relaxation, τ_ω, is defined as the longest relaxation time for subsequences of length 2 or larger. In Appendices A and B, we provide calculations to justify the expression of τ_l and τ_ω given in Table I.

In our simulations, we have chosen τ_ω in order to construct a dimensionless time $\hat{t} = t/\tau_\omega$. We start with an initial population of molecules, and then use the Gillespie algorithm to generate a trajectory through the space of compositions. In order to evaluate the various contributions to the total entropy production introduced in (41).

In Fig. 3, simulation results for the two contributions to the system entropy per chain, namely, Δs_L and Δs_ω are shown as a function of time. The total number of chains is either $N = 32$ or $N = 2042$, and the initial condition has an equal amount of 000 and 111 chains. The figure shows that at time $\hat{t} = 3$, the system has reached equilibrium. This equilibrium, however, differs from the macroscopic equilibrium, corresponding to the dashed lines, when N is small. When the system size is sufficiently large (in our simulation: $N = 2048$), there is a good

TABLE I. Expressions of the various relaxation times: τ_r waiting time for a reaction to occur, τ_l relaxation time of the length, τ_ω relaxation time of the sequence.

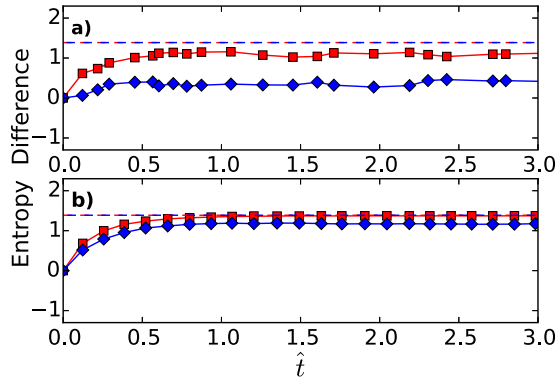| Reaction	τ_r	τ_l	τ_ω
Attack-exchange	$\frac{1}{kN(M-N)}$	$\frac{1}{kM}$	$\frac{1}{kN}$
Chain-exchange	$\frac{2}{k(M-N)^2}$	$\frac{1}{k(M-N)}$	$\frac{1}{k(M-N)}$

FIG. 3. Difference in length and sequence entropy per chain: Δs_L (red square) and Δs_ω (blue diamond), as function of time $\hat{t}$ for chain-exchange. The number of polymers are (a) $N = 32$ or (b) $N = 2048$. The initial composition consists of sequences 000 and 111, in equal abundance. The entropy differences in the thermodynamic limit, ΔS_L^{eq} and ΔS_ω^{eq} are shown as dashed lines.

agreement with the values of Eq. (26),

$$\Delta \mathcal{S}_L^{eq} = \sum_l \left[Y_l^I \ln Y_l^I - Y_l^{eq} \ln Y_l^{eq} \right],$$

$$\Delta \mathcal{S}_\omega^{eq} = \sum_{l,\Omega} \left[Y_l^I U_{l,\Omega}^I \ln U_{l,\Omega}^I - Y_l^{eq} U_{l,\Omega}^{eq} \ln U_{l,\Omega}^{eq} \right]. \quad (53)$$

As can be seen in Fig. 3, the weighted sequence disorder Δs_ω differs more from its macroscopic expression than the length disorder Δs_L for small values of N. The reason is that many sequences are not present or not sufficiently abundant in that case, which explains the lack of convergence to the macroscopic limit.

D. Entropic exchange induced by partially equilibrated sequences

We now focus on a case where the initial condition of the system is out of equilibrium for the length distribution but in a state of *partial* equilibrium for the sequence (given the chosen initial length). In that case, the weighted sequence entropy starts initially at its maximal value, while the length entropy is not maximum.

This case is illustrated in Fig. 4. Since we plot the entropy difference with respect to the final equilibrium value, the
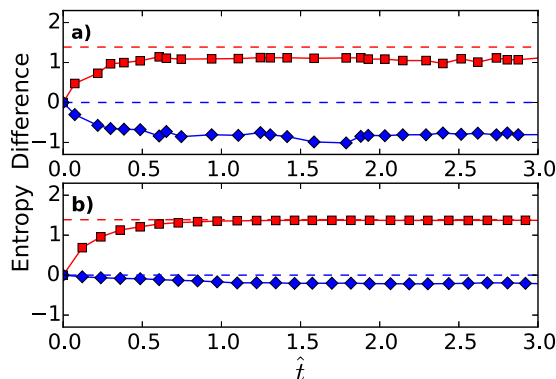

FIG. 4. Idem as in Fig. 3, except that the initial composition consists of 8 sequences of length 3 that can be made with two monomer types in equal amount.
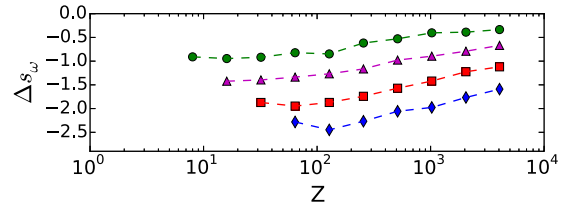

FIG. 5. Difference in sequence entropy per chain after relaxation, averaged over a time window of $30\tau_\omega$, as function of the number of chains N. The initial condition contains the 2^{l_A} sequences that exist for a given length l_A, with: $l_A = 3$ (green circle), $l_A = 4$ (purple diamond), $l_A = 5$ (red square), and $l_A = 6$ (blue diamond). In the thermodynamic limit, we have $\Delta S_\omega^{eq} = 0$.

difference of sequence entropy starts at zero and then becomes negative. In the figure, the macroscopic limit of that quantity shown as the dashed blue line is zero. This is easy to verify. Indeed, if the sequence is relaxed from the start,

$$\Delta \mathcal{S}_\omega^{eq} = \sum_{l,\Omega} [Y_l^I - Y_l^{eq}] l \ln m$$

$$= \left[\frac{M}{N} - \frac{M}{N} \right] \ln m = 0. \quad (54)$$

In the course of the simulation, the length distribution broadens. For the short polymers, there will typically be enough polymers to have a complete set of all the sequences for that length. However, for longer polymers, many sequences will be absent. As a result, the sequence entropy cannot reach its macroscopic limit.

In any case, the negative contribution of the sequence entropy is offset by that of the length entropy in agreement with the second law that imposes that the sum of the two terms be positive. It is important to point out that this finite size effect only exists for specific choices of initial conditions and disappears in the thermodynamic limit when $N \to \infty$.

We have studied the dependence of this effect for various polymer lengths as shown in Fig. 5. In this figure, we have chosen the initial condition of the system to be an ensemble of polymers of the same length with $Y_l = \delta_l^{l_A}$ with a complete set of all possible 2^{l_A} sequences, uniformly distributed. We then evaluate Δs_ω by averaging over a time window of length $30\tau_\omega$, after at least $3\tau_\omega$ have elapsed. The figure shows that this time averaged Δs_ω decreases with increasing l_A at a fixed number of chains. This is compatible with the fact that Δs_ω is largely controlled by the weighted sequence entropy of the initial state. As l_A increases, so does the number of configurations in the initial state, and therefore also its sequence entropy.

VI. SIMULATIONS WITH ENERGETICALLY NON-NEUTRAL REACTIONS

In this section, we introduce a simple example of an energy landscape. We assume that there is a certain local energy function, dependent on the nature of the bonds between the nearest neighboring monomers in a sequence. We denote with $\tilde{n}_\omega$ the total number of bonds ω, which are present among all the polymers of the system. This number is

$$\tilde{n}_\omega = \sum_{\omega_A,\omega_B} n_{\omega_A \omega \omega_B}. \quad (55)$$

When only two monomer types are present, the only relevant exchange reaction at the level of subsequences is

$$\omega_A 00 \omega_B + \omega_C 11 \omega_D \rightleftharpoons \omega_C 10 \omega_B + \omega_A 01 \omega_D, \quad (56)$$
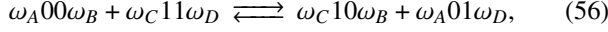
since the other reactions do not change bond composition. Let us introduce the standard chemical potential of the various bonds: $\tilde{\mu}_{00}^\circ$, $\tilde{\mu}_{01}^\circ$, $\tilde{\mu}_{10}^\circ$ and $\tilde{\mu}_{11}^\circ$. Then the forward rate of reaction (56) is $k^+ \sim \exp\left(-\beta(\tilde{\mu}_{00}^\circ + \tilde{\mu}_{11}^\circ)\right)$ while the backward rate is $k^- \sim \exp\left(-\beta(\tilde{\mu}_{01}^\circ + \tilde{\mu}_{10}^\circ)\right)$. The detailed balance condition imposes

$$\frac{\tilde{n}_{00}^{eq} \tilde{n}_{11}^{eq}}{\tilde{n}_{01}^{eq} \tilde{n}_{10}^{eq}} = \frac{k^-}{k^+} = \exp\left(-\beta \Delta \tilde{\mu}^\circ\right), \quad (57)$$

in terms of the standard chemical potential change $\Delta \tilde{\mu}^\circ = \tilde{\mu}_{01}^\circ + \tilde{\mu}_{10}^\circ - \tilde{\mu}_{00}^\circ - \tilde{\mu}_{11}^\circ$. In practice, the reaction (56) can only occur if the two reacting subsequences are present on different polymer chains. In order to simplify the modeling, we use a mean-field approximation, which corresponds to assuming that any subsequence can react with any other subsequence independently of the chain that carries them.

Using (57), we can find $\tilde{n}_\omega^{eq}$ at equilibrium and compare it to its initial values $\tilde{n}_\omega^I$. In order to evaluate the heat, we assume that there is no change of internal entropy during a recombination reaction, which means that $\Delta \tilde{s}^\circ = 0$ at all times. As a result, on long times, the stochastic heat defined in Eq. (34), $q(t \to \infty)$ equals the difference in standard chemical potential,

$$q(t \to \infty) = (\tilde{n}_{00}^{eq} - \tilde{n}_{00}^I)\Delta \tilde{\mu}^\circ. \quad (58)$$

We consider two kinds of initial sequences: (a) of the form 0 101 010 101 and 1 010 101 010 in equal abundance, and (b) of the form 0 000 000 000 and 1 111 111 111 in equal abundance. In both cases, we took $\Delta \tilde{\mu}^\circ = -2k_BT$. Therefore, (a) is high in energy because it is rich in 01 and 10 bonds and (b) is low in energy since it is rich in 00 and 11 bonds. As a result, we expect $q(t \to \infty)$ to be negative for case (a) and positive for case (b). We will now proceed to find the equilibrium distributions, in order to calculate the entropy changes for $N \to \infty$.

Let us assume a symmetric initial condition, in the relative amount of subsequences 00 and 11, including terminal and initial positions. Since the only relevant reaction is given by Eq. (56), this symmetry will persist and we will have $\tilde{n}_{00} = \tilde{n}_{11}$ and $\tilde{n}_{01} = \tilde{n}_{10}$ at all times. As a result, Eq. (57) simplifies into

$$\frac{\tilde{n}_{00}^{eq}}{\tilde{n}_{01}^{eq}} = \exp\left(\frac{-\beta \Delta \tilde{\mu}^\circ}{2}\right). \quad (59)$$

The free energy of the system can be written in terms of: (i) entropy of the length distribution, (ii) standard free energy of the subsequences, (iii) entropy of the subsequence distribution. Since (i) is not coupled to (ii) and (iii), we can maximize (i) independently. Consequently, we obtain the same length distribution as in the energetically neutral case: (50). For less symmetric cases or more complex energy landscapes, Y_l^{eq} should be modified.

An explicit expression of the equilibrium sequence distribution for given length: $U_{l,\Omega}^{eq}$ can be found from the following argument. A given sequence Ω has an energy e_Ω corresponding to its bond composition. We define n_B as the number of
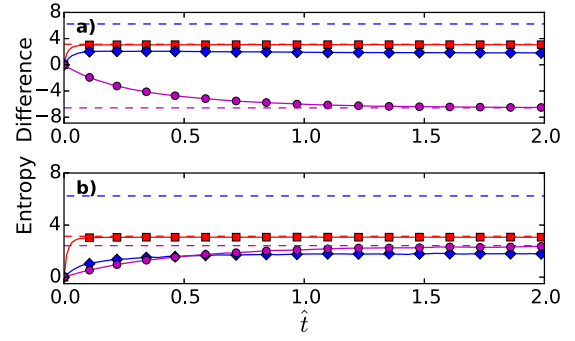

FIG. 6. Difference in length entropy Δs_L (red square), sequence entropy Δs_ω (blue diamond), and heat per chain q/N (purple circle) as function of time $\hat{t}$ for a variant of chain-exchange dynamics with an energy function dependent on neighboring bonds. We start with $N = 2048$ polymers of sequence: (a) 0 000 000 000 and 1 111 111 111;(b) 0 101 010 101 and 1 010 101 010. Dashed lines represent thermodynamic limits as in previous figures.

bonds of the type 00 and 11 in Ω. Therefore: $e_\Omega = n_B \Delta \tilde{\mu}^\circ / 2$. There are $2\binom{l-1}{n_B}$ species of length l with n_B of such bonds. We thus find for $U_{l,\Omega}^{eq}$,

$$U_{l,\Omega}^{eq} = \frac{\exp(-\beta e_\Omega)}{\sum_{n_B=0}^{l-1} 2\binom{l-1}{n_B} \exp\left(-\frac{\beta n_B \Delta \tilde{\mu}^\circ}{2}\right)}$$
$$= \frac{\exp(-\beta e_\Omega)}{2\left(1 + \exp\left(-\frac{\beta \Delta \tilde{\mu}^\circ}{2}\right)\right)^{l-1}}. \quad (60)$$

To perform simulations, we use the Gillespie scheme with an energy-dependent rejection Monte Carlo step, where rejections lead to the repetition of this selection until a next reaction is accepted. With this energy landscape, the sequence relaxation time becomes (see Appendix A for a derivation of that result)

$$\tau_\omega = \frac{\exp\left(-\frac{\beta \Delta \tilde{\mu}^\circ}{2}\right)}{k(M - N)}. \quad (61)$$

This calculation shows that the modification of the characteristic time of relaxation of the sequence is the main effect of introducing energy landscape, at least in this simple model.

In Fig. 6(a), heat is liberated, as additional bonds of type 00 and 11 are formed. In Fig. 6(b), the system takes up heat from the environment. In all cases, we have: $\Delta s_L + \Delta s_\omega - \beta q/N \geq 0$. In this symmetric example, the energy landscape only affects the sequence, not the length. When sequence complementarity or secondary structure is considered, this may no longer be the case.

VII. CONCLUSION

In this work, we have studied the relaxation of a pool of information-carrying polymers kept in a closed system but dynamically evolving under the action of reversible recombination reactions. We have focused on two types of recombination reactions, namely, attack-exchange and chain-exchange because they are simple, energetically neutral, robust, and potentially relevant for prebiotic chemistry since they do not require catalysts. We have developed a stochastic thermodynamic framework to analyze the dynamic evolution of thermodynamic quantities such as heat or entropy under such reactions.

Inspired by the work of Andrieux and Gaspard,[10] we have introduced in Eq. (26) a splitting of the entropy production into three contributions: the length disorder, the sequence disorder weighted by the length distribution, and the standard free energy change. This key result indicates that for finite systems a coupling exists between weighted sequence disorder and length disorder. Thus, we find that for some choice of initial conditions, the weighted sequence disorder can be decreased at the expense of an increase of length disorder; a finite size effect that however disappears in the thermodynamic limit.

In the context of prebiotic systems, an important question is whether or not recombination reactions can lead to the formation of long, catalytically active polymers. For simple energy landscapes, length distributions become exponential distributions, which only yield a small amount of large polymers. In order to obtain non-exponential distributions, at least one of the following ingredients is needed (i) energy landscapes favoring long species, (ii) time dependent forcing, and (iii) exchange with an environment.

We have studied here the effect of an energy landscape (i) by assuming that the energy lies only in neighboring bonds. Despite the simplicity of that assumption, we have observed that it leads to a modification of the characteristic time of relaxation of the sequence. Clearly, this is one of the simplest cases and energy constraints can affect the dynamics in more complex ways due to secondary and tertiary structure of nucleic acids. Often, the secondary and tertiary structure of polymers affects their collective interactions.[39]

Many possibilities exist concerning (ii), whereby a time dependent forcing in the bulk rate constants or in boundary conditions can affect the kinetics of polymerization such as in day-night models of polymerization.[40] Concerning (iii), one way to describe the coupling of a system to an environment is to introduce chemostats that impose that the concentration of certain polymers be fixed. We have found in previous work that such models have a rich dynamics even for polymers that have no sequences.[41] Such an approach based on Stochastic Thermodynamics was extended for general chemical networks in Ref. 20. We plan to explore in future work such an approach to polymers that have a sequence. Finally, another interesting research direction for (iii) concerns the exchange with a structured environment, which can take the form of compartments as in Ref. 42 or more generally any element with a scaffolding function.

ACKNOWLEDGMENTS

APPENDIX A: SEQUENCE RELAXATION DYNAMICS

This section contains a derivation of the length and sequence characteristic times that have been introduced in

Table I. We start with the relaxation time of the sequence, which has been defined as the characteristic time of randomization of subsequences of length two as a result of the reaction (56). We consider here only the chain-exchange reaction.

Let us assume that the initial condition is symmetric with respect to the content of 0 and 1 monomers in the pool. As a result, this symmetry will remain at all times, and we can introduce x and y variables such that $\tilde{n}_{00} = \tilde{n}_{11} = x$ and $\tilde{n}_{01} = \tilde{n}_{10} = y$. In the mean-field approximation, the evolution of equations of these variables are

$$\frac{dx}{dt} = k^- y^2 - k^+ x^2,$$
$$\frac{dy}{dt} = k^+ x^2 - k^- y^2, \tag{A1}$$

where k^+ is a forward rate and k^- a backward rate. By summing the two equations above, one recovers the conservation law that the sum of x and y is constant. The constant is fixed by the initial number of bonds: $2x + 2y = M - N$. Therefore, we end up with the equation

$$\frac{dx}{dt} = k^- \left(\frac{M-N}{2} - x \right)^2 - k^+ x^2. \tag{A2}$$

For neutral reactions, $k^+ = k^- = k$, the equation simplifies into

$$\frac{dx}{dt} = -k \left[(M-N)x - \left(\frac{M-N}{2} \right)^2 \right]. \tag{A3}$$

This linear ODE has a simple exponential as solution with the characteristic relaxation time $\tau_\omega = 1/k(M-N)$, which was given in Table I.

Let us now extend the above results to the case that transitions are affected by an energy landscape. We start with the detailed balance condition: $k^+ = k^- \exp\left(-\beta \Delta \tilde{\mu}^\circ\right)$. We now go back to Eq. (A2) when $k^- \neq k^+$. We obtain a nonlinear ODE of the form

$$\frac{dx}{dt} = ax^2 + bx + c. \tag{A4}$$

With a, b, and c constants, given by

$$a = k^- - k^+,$$
$$b = k^-(M-N),$$
$$c = k^- \left(\frac{M-N}{2} \right)^2, \tag{A5}$$

we note that $\sqrt{b^2 - 4ac} = \sqrt{k^+ k^-}(M-N) > 0$. Therefore, we can make use of the integral

$$\int_0^t dt = \int_{x(0)}^{x(t)} \frac{dx}{ax^2 + bx + c}$$
$$= \frac{-2}{\sqrt{b^2 - 4ac}} \tanh^{-1}\left(\frac{2ax(t) + b}{\sqrt{b^2 - 4ac}} \right) + C. \tag{A6}$$

Therefore, the solution is of the form

$$x(t) \propto \tanh\left[\frac{-\sqrt{b^2 - 4ac}}{2}(t - C) \right] + D, \tag{A7}$$

where C and D are constants. As $\tanh(t) = (1 - \exp(-2t))/(1 + \exp(-2t))$, we can identify $1/\sqrt{b^2 - 4ac}$ as a characteristic sequence relaxation time τ_ω equal to

$$\tau_\omega = \frac{\exp\left(-\frac{\beta\Delta\tilde{\mu}^\circ}{2}\right)}{k^+(M-N)}, \tag{A8}$$

which is precisely Eq. (61) of the main text.

APPENDIX B: LENGTH RELAXATION DYNAMICS

We now derive the characteristic time of length relaxation, first for chain-exchange, and then for attack-exchange. The amount of species of length l: N_l evolves according to

$$
\begin{aligned}
\frac{dN_l}{dt} &= k \sum_{l_A+l_B=l} \sum_{l_C,l_D}^\infty [N_{l_A+l_D}N_{l_C+l_B} - N_{l_A+l_B}N_{l_C+l_D}] \\
&= k \sum_{l_C,l_D}^\infty \sum_{l_B=1}^{l-1} N_{l_D+l-l_B}N_{l_C+l_B} - k(l-1)N_l \sum_{l_x=2}^\infty (l_x-1)N_{l_x} \\
&= k \sum_{l_B=1}^{l-1} \left(N - \sum_{l_x=2}^{l-l_B} N_{l_x}\right)\left(N - \sum_{l_y=2}^{l_B} N_{l_y}\right) - k(l-1)(M-N)N_l.
\end{aligned} \tag{B1}
$$

Therefore, the homogeneous equation takes the form

$$\frac{dN_l}{dt} = k(l-1)N^2 - k(l-1)(M-N)N_l, \tag{B2}$$

which admits the solution

$$N_l = \frac{N^2}{M-N} + \left(N_l^I - \frac{N^2}{M-N}\right)\exp[-k(l-1)(M-N)t]. \tag{B3}$$

We have introduced N_l^I as the initial value of $N_l(t)$. Note that the exponential in the homogeneous solution is proportional to $l-1$, while the highest possible order in the particular solution, arising from terms such as $N_{l-l_B}N_{l_B} \propto \exp(-k(l-l_B-1)(M-N)t)\exp(-k(l_B-1)(M-N)t) = \exp(-k(l-2)(M-N)t)$, is $l-2$. We therefore have no resonant terms for any N_l, and we can expect a solution of Eq. (B1) of the form

$$N_l = A_{0,l} + \sum_{n=2}^l A_{n,l}\exp[-k(n-1)(M-N)t], \tag{B4}$$

where $A_{0,l}$ and $A_{n,l}$ are constants depending on initial concentrations of all species. This expression confirms that the slowest relaxation time of the length for chain-exchange reaction is $\tau_l = 1/(k(M-N))$ as given in Table I.

For attack-exchange, the kinetic equation for N_l is

$$
\begin{aligned}
\frac{dN_l}{dt} &= k \sum_{l_A,l_B=1}^\infty [N_{l_A}N_{l+l_B} - N_{l_A+l_B}N_l] \\
&\quad + k \sum_{l_A} \sum_{l_B=1}^{l-1} [N_{l_A}N_{l_C+l-l_A} - N_l N_{l_C}] \\
&= k\Big[N(N - \sum_{l_B=1}^{l-1} N_{l_B}) + \sum_{l_B=1}^{l-1} N_{l-l_B}(N - \sum_{l_x=1}^{l_B}) \\
&\quad - (M+N(l-1))N_l\Big].
\end{aligned} \tag{B5}
$$

Upon solving the homogeneous equations, the general solution for every N_l can be written as

$$N_l = A_{0,l} + \sum_{n=1}^l A_{n,l} \quad \exp[-k(M+N(l-1))t]. \tag{B6}$$

For which the longest relaxation time is: $\tau_l = \frac{1}{kM}$.

[1] A. Murugan, Z. Zeravcic, M. P. Brenner, and S. Leibler, Proc. Natl. Acad. Sci. U. S. A. **112**, 54 (2015).
[2] G. Gines, A. S. Zadorin, J.-C. Galas, T. Fujii, A. Estevez-Torres, and Y. Rondelez, Nat. Nanotechnol. **12**, 351 (2017).
[3] D. Y. Zhang, A. J. Tuberfield, B. Yurke, and E. Winfree, Science **318**, 1121 (2007).
[4] N. Badi and J.-F. Lutz, Chem. Soc. Rev. **38**, 3383 (2009).
[5] J.-F. Lutz, M. Ouchi, D. Liu, and M. Sawamoto, Science **341**, 1238149 (2013).
[6] P. J. Flory, J. Chem. Phys. **12**, 425 (1944).
[7] P. J. Blatz and A. V. Tobolsky, J. Chem. Phys. **49**, 77 (1945).
[8] T. L. Hill, *Free Energy Transduction and Biochemical Cycle Kinetics* (Springer, 1989).
[9] F. Oosawa and S. Asakura, *Thermodynamics of the Polymerization of Protein* (Academic Press, 1975).
[10] D. Andrieux and P. Gaspard, Proc. Natl. Acad. Sci. U. S. A. **105**, 9516 (2008).
[11] P. Gaspard and D. Andrieux, J. Chem. Phys. **141**, 044908 (2014).
[12] P. Gaspard, Phys. Rev. E **93**, 042420 (2016).
[13] P. Gaspard, Phys. Rev. Lett. **117**, 238101 (2016).
[14] T. E. Ouldridge and P. Rein ten Wolde, Phys. Rev. Lett. **118**, 158103 (2017).
[15] Ö. Kartal, S. Mahlow, A. Skupin, and O. Ebenhöh, Mol. Syst. Biol. **7**, 542 (2011).
[16] S. Lahiri, Y. Wang, M. Esposito, and D. Lacoste, New J. Phys. **17**, 085008 (2015).
[17] U. Seifert, Rep. Prog. Phys. **75**, 126001 (2012).
[18] Y. D. Decker, Phys. A **428**, 178 (2015).
[19] M. Esposito and C. V. den Broeck, Phys. A **418**, 6 (2015).
[20] R. Rao and M. Esposito, Phys. Rev. X **6**, 041064 (2016).
[21] P. G. Higgs and N. Lehman, Nat. Rev. Genet. **16**, 7 (2015).
[22] S. Y. Nechaev, A. V. Lutay, V. V. Vlassov, and M. A. Zenkova, Int. J. Mol. Sci. **10**, 1788 (2009).
[23] N. Lehman, Chem. Biodiversity **5**, 1707 (2008).
[24] J. Derr, M. Manapat, S. Rajamani, K. Leu, R. Xulvi-Brunet, I. Joseph, M. Nowak, and I. Chen, Nucleic Acids Res. **40**, 4711 (2012).
[25] N. Lehman, J. Mol. Evol. **56**, 770 (2003).
[26] Y. Lee and D. C. Rio, Annu. Rev. Biochem. **84**, 291 (2015).
[27] Y. Shao and S. B. Kent, Chem. Biol. **4**, 187 (1997).
[28] N. Vaidya, M. L. Manapat, I. A. Chen, R. Xulvi-Brunet, E. J. Hayden, and N. Lehman, Nature **491**, 72 (2012).
[29] C. Bonfio, L. Valer, S. Scintilla, S. Shah, D. J. Evans, L. Jin, J. W. Szostak, D. D. Sasselov, J. D. Sutherland, and S. S. Mansy, "UV-light-driven prebiotic synthesis of iron–sulfur clusters," Nat. Chem. (published online, 2017).
[30] A. H. Hoveyda and A. R. Zhugralin, Nature **450**, 243 (2007).
[31] D. Montarnal, M. Capelot, F. Tournilhac, and L. Leibler, Science **334**, 965 (2011).
[32] M. Capelot, D. Montarnal, F. Tournilhac, and L. Leibler, J. Am. Chem. Soc. **134**, 7664 (2012).
[33] W. Denissen, G. Rivero, R. Nicolay, L. Leibler, J. M. Winne, and F. E. Du Prez, Adv. Funct. Mater. **25**, 2451 (2015).
[34] W. Denissen, J. M. Winne, and F. E. Du Prez, Chem. Sci. **7**, 30 (2016).
[35] D. Kondepudi and I. Prigogine, *Modern Thermodynamics: From Heat Engines to Dissipative Structures* (Wiley, New York, 1998).
[36] P. Gaspard, J. Chem. Phys. **120**, 8898 (2004).
[37] T. Schmiedl, T. Speck, and U. Seifert, J. Stat. Phys. **128**, 77 (2007).
[38] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[39] B. Obermayer, H. Krammer, D. Braun, and U. Gerland, Phys. Rev. Lett. **107**, 018101 (2011).
[40] A. V. Tkachenko and S. Maslov, J. Chem. Phys. **143**, 045102 (2015).
[41] R. Rao, D. Lacoste, and M. Esposito, J. Chem. Phys. **143**, 244903 (2015).
[42] S. Matsumura, Á. Kun, M. Ryckelynck, F. Coldren, A. Szilágyi, F. Jossinet, C. Rick, P. Nghe, E. Szathmáry, and A. D. Griffiths, Science **354**, 1293 (2016).