# Nonequilibrium generation of information in copolymerization processes

**David Andrieux** [*][†] **and Pierre Gaspard** [†]

[†]Center for Nonlinear Phenomena and Complex Systems, Université Libre de Bruxelles, Code Postal 231, Campus Plaine, B-1050 Brussels, Belgium

**We consider general fluctuating copolymerization processes, with or without underlying templates. The dissipation associated with these nonequilibrium processes turns out to be closely related to the information generated. This shows in particular how information acquisition results from the interplay between stored patterns and dynamical evolution in nonequilibrium environments. In addition, we apply these results to the process of DNA replication.**

DNA replication │ entropy production │ nonequilibrium fluctuations │ self-organization

The origin of biological information is one of the major challenges for our understanding of the living organisms. Since the discovery of DNA, the biochemical support for the storage of genetic information is known. DNA is a copolymer which keeps the memory of information on the living organism in its structure. This molecular structure is stable at ambient temperatures by the binding energy between the nucleotides, allowing the heredity of genetic information across generations. As observed *in vitro* in evolution experiments on RNA and viruses [1, 2], the processing of biological information can be discussed in terms of the dynamics of populations associated with the different possible genetic sequences, the populations evolving by replications and mutations into quasi-species [3, 4]. Such population dynamics are nonequilibrium processes where dissipation is compensated by energy supply and the entropy produced by dissipation is evacuated to the environment of these open systems. However, this view relies on macroscopic concepts such as the population sizes, which are largely separated from the nanoscale of the genetic sequences. Moreover, observations reveal that biological systems have structures and functions at every scale down to the molecular level and the understanding of their origin is a challenge.

Actually, the information in DNA copolymers is processed and replicated by mechanisms taking place at the molecular level in the presence of thermal fluctuations. These fluctuations are due to the random motion of the atoms and molecules composing DNA, the transcription or replication machinery, and their environment. In this regard, biological information processing is ruled by the statistical laws of motion and thermodynamics. At the thermodynamic equilibrium, the principle of detailed balance implies that no information can be spontaneously processed or generated because each random motion is statistically balanced by the corresponding reversed motion. Therefore, equilibrium is the stage of erratic motion where information generation is highly improbable.

Recently, it has been shown that nonequilibrium fluctuating systems present a time asymmetry in which the typical random paths followed by the system during its time evolution turn out to be more probable than their time reversal [5, 6, 7, 8]. The remarkable result is that this temporal ordering of nonequilibrium fluctuations is the consequence of the second law of thermodynamics. This phenomenon explains that dynamical order can be naturally generated in molecular motion under nonequilibrium conditions.

The goal of this paper is thus to show that the implication of this nonequilibrium temporal ordering in copolymerization processes is the generation of information. Indeed, copolymerization can store in space the dynamical order which is generated in time by the nonequilibrium processes, establishing a coupling from time back to the spatial support of information. The possibility of such transfers of information has already been envisaged at the macroscopic level [9] and is here shown to arise at the molecular level. With this aim, we study copolymerization processes with or without the presence of an underlying template. Such processes have the ability to store information in the copolymer pattern, which in turn influences back the dynamics of the system. Information generation can thus be understood as a natural consequence of driving out of equilibrium systems with the ability to store and retrieve molecular information.

We first explain the tradeoff between dissipation and information in copolymerization processes in the framework of nonequilibrium statistical thermodynamics. We then proceed with the illustration of these results in the case of DNA replication.

## Dissipation-Information Tradeoff

Copolymerization processes can be described at the macroscopic level in terms of the population sizes, i.e., the chemical concentrations of the monomers, dimers, trimers, etc... In this macroscopic description, the rate of dissipation, in particular, due to metabolism can be estimated by the thermodynamic entropy production given in terms of the chemical concentrations [10]. Since we are interested in the molecular structure generated during copolymerization, we adopt a mesoscopic description at the scale of a single copolymer chain.

We first consider a copolymerization process without template and with $M$ different monomers (see Fig. 1a). We denote by $\omega$ a copolymer chain of finite length surrounded by a solution containing monomers $m$ at given concentrations. The solution surrounding each single copolymer chain is assumed to be sufficiently large to play the role of a reservoir which is continually replenished with new monomers so as to keep their concentrations fixed. In this situation, an arbitrary chain $\omega$ can grow by adding a monomer $m$: $\omega + m \rightarrow \omega m$. This copolymerization step is driven by the free-enthalpy change of the reaction. If this latter is high enough, the monomers attach to but never detach from the chain and the process is totally irreversible. For instance, if the binding rates $k_{+m}$ only depend on the monomer $m$ added to the chain, the growth proceeds by the random addition of

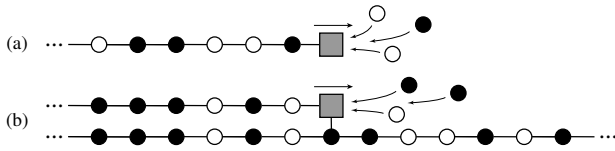[*]To whom correspondence should be addressed. E-mail: dandrieu@ulb.ac.be

**Fig. 1.** Schematic representations of a copolymerization process without a template (a) and a copolymerization process with a template (b). The circles depict the monomers and the square the catalyst of polymerization.

monomers at the velocity $v = \sum_m k_{+m}$. The insertion probabilities are here given by $\eta_m = k_{+m}/\sum_n k_{+n}$, showing that the chain may be disordered due to the multiplicity of the monomer species composing the copolymer. However, such totally irreversible processes could only happen in the extreme case where the copolymerization is driven arbitrarily far from equilibrium. In general, the free-enthalpy differences between the reactants $\omega + m$ and the products $\omega m$ are finite and reversed reactions are possible. Accordingly, the chain $\omega$ may not only grow by the addition of monomers, but also shorten by removing the last monomer $m$ of the actual chain: $\omega + m \rightleftharpoons \omega m$. The closer to equilibrium the process is, the more probable the reversed reactive events are. Consequently, the configurational disorder of the chain can be larger if it is grown under low driving forces in the vicinity of the thermodynamic equilibrium.

The general framework to study such processes is provided by the stochastic theory of nonequilibrium processes [11, 12]. At the mesoscopic level, many phenomena such as nonequilibrium reactions are successfully described in terms of continuous-time random processes. In this framework, the probability $P_t(\omega)$ to find the copolymer chain $\omega$ at time $t$ is ruled by the evolution equation:

$$\frac{dP_t(\omega)}{dt} = \sum_{\omega'} \left[ P_t(\omega') W(\omega', \omega) - P_t(\omega) W(\omega, \omega') \right] , \qquad (1)$$

where the coefficients $W(\omega, \omega')$ denote the rates of the transitions $\omega \to \omega'$ [11, 12]. For many processes at fixed pressure and temperature $T$, the ratio of forward to backward transition rates can be expressed in terms of the free enthalpy $G(\omega)$ of a single copolymer chain $\omega$ surrounded by the solution, according to

$$\frac{W(\omega, \omega')}{W(\omega', \omega)} = \exp \frac{G(\omega) - G(\omega')}{k_B T} , \qquad (2)$$

where $k_B$ is Boltzmann's constant [13]. We notice that the transition rates $W(\omega, \omega')$ as well as the free enthalpies $G(\omega)$ will in general depend on the concentrations of monomers $m$ in the solution surrounding the chain $\omega$ [11, 12, 13].

A key concept for our purposes is the thermodynamic entropy of the system:

$$\mathcal{S}(t) = \sum_\omega P_t(\omega) S(\omega) - k_B \sum_\omega P_t(\omega) \ln P_t(\omega) , \qquad (3)$$

where $S(\omega)$ is the entropy of the copolymer chain $\omega$ in solution [14]. The entropy (Eq. **3**) varies in time with the probability distribution ruled by the evolution equation (Eq. **1**). It is known that these variations $d\mathcal{S}/dt$ are due to both the exchange of entropy with the environment $d_e\mathcal{S}/dt$ and the internal production of entropy

$$\frac{d_i\mathcal{S}}{dt} = \frac{d\mathcal{S}}{dt} - \frac{d_e\mathcal{S}}{dt} \geq 0 , \qquad (4)$$

which is always non negative in agreement with the second law of thermodynamics [11, 12, 15]. The exchange of entropy is given by

$$\frac{d_e\mathcal{S}}{dt} = \frac{1}{T} \frac{d\langle H \rangle_t}{dt} \qquad (5)$$

in terms of the time variation of the mean enthalpy of the system: $\langle H \rangle_t = \sum_\omega P_t(\omega) H(\omega)$ [14, 15]. The enthalpy $H(\omega)$ of the copolymer $\omega$ is related to the corresponding free enthalpy and entropy by $G(\omega) = H(\omega) - TS(\omega)$. We notice that the exchange of entropy during the time interval $\Delta t$ is equal to the heat of copolymerization $\Delta\langle H \rangle$ over the temperature $T$ according to $\Delta_e\mathcal{S} = \Delta\langle H \rangle/T$.

It is assumed that, for fixed concentrations of monomers, the growth process can go on without limit and possibly reach a regime where the synthesized copolymer has stationary statistical properties [16, 17]. In such a stationary regime, the mean length of the chains grows at some fixed average velocity

$$v = \frac{d\langle l \rangle_t}{dt} . \qquad (6)$$

The mean length is evaluated as $\langle l \rangle_t = \sum_l p_t(l) \times l$ in terms of the probability distribution $p_t(l)$ of the chain lengths $l$ at time $t$. The statistical composition of the copolymer in the stationary regime is described by the probability distribution $\mu_l(\omega)$ to have a chain $\omega$ provided that its length is $l$. This distribution is normalized as $\sum_\omega \mu_l(\omega) = 1$, where the sum is restricted to all the chains $\omega$ of length $l$. Accordingly, the probability to have a certain chain $\omega$ at time $t$ becomes

$$P_t(\omega) = p_t(l) \, \mu_l(\omega) , \qquad (7)$$

in the stationary regime [16, 17].

The entropy (Eq. **3**) can thus be rewritten as

$$\mathcal{S}(t) = \sum_{l,\omega} p_t(l) \, \mu_l(\omega) \, S(\omega) \qquad (8)$$

$$-k_B \sum_{l,\omega} p_t(l) \, \mu_l(\omega) \ln \mu_l(\omega) - k_B \sum_l p_t(l) \ln p_t(l) .$$

The first term is the contribution from the thermodynamic entropy $S(\omega)$ of the copolymer $\omega$ in its environment. In the stationary regime, we may introduce the mean entropy per monomer as

$$s = \lim_{l \to \infty} \frac{1}{l} \sum_\omega \mu_l(\omega) \, S(\omega) \qquad (9)$$

so that the first term becomes $\langle l \rangle_t \times s$, in the long-time limit.

The second term in Eq. **8** can be expressed in terms of the disorder in the composition of the copolymer. This disorder can be characterized by the Shannon entropy per monomer defined as

$$D(\text{polymer}) = \lim_{l \to \infty} -\frac{1}{l} \sum_\omega \mu_l(\omega) \ln \mu_l(\omega) . \qquad (10)$$

The disorder takes a value in the range $0 \leq D \leq \ln M$ limited by the number of different monomers composing the copolymer. In the stationary regime, the second term is thus controlled by the disorder in the chain since

$$-\sum_{l,\omega} p_t(l)\mu_l(\omega) \ln \mu_l(\omega) \qquad (11)$$

$$= \sum_l p_t(l) \, l \left[ -\frac{1}{l} \sum_\omega \mu_l(\omega) \ln \mu_l(\omega) \right] \simeq \langle l \rangle_t \, D(\text{polymer})$$

at long times.

The third term in Eq. **8** is the Shannon entropy of the distribution of lengths. This distribution broadens in time and its entropy typically increases logarithmically in time. Accordingly, this term does not contribute to the variation rate of the entropy because $\lim_{t \to \infty}(-1/t) \sum_l p_t(l) \ln p_t(l) = 0$.

Hence, we find that the time variation of the entropy is given in the stationary regime by

$$\frac{d\mathcal{S}}{dt} = \lim_{t\to\infty} \frac{\mathcal{S}(t)}{t} = v\,s + v\,D(\text{polymer})\,, \qquad \textbf{(12)}$$

with the growth velocity (Eq. **6**) and physical units where $k_{\mathrm{B}} = 1$ used for the simplicity of notations. On the other hand, the entropy exchange rate (Eq. **5**) is equal to

$$\frac{d_{\mathrm{e}}\mathcal{S}}{dt} = v\,\frac{h}{T}\,, \qquad \textbf{(13)}$$

where the mean enthalpy per monomer $h$ has a definition similar to Eq. **9**. Finally, the entropy production (Eq. **4**) reads

$$\frac{d_{\mathrm{i}}\mathcal{S}}{dt} = v\,A \geq 0\,, \qquad \textbf{(14)}$$

with the growth velocity (Eq. **6**) and the thermodynamic force or affinity per monomer

$$A = -\frac{g}{T} + D(\text{polymer}) = \varepsilon + D(\text{polymer})\,, \qquad \textbf{(15)}$$

where $g = h - Ts$ is the free enthalpy per monomer and $\varepsilon = -g/T$ is the mean driving force. We note that these results do not depend on the specific choice of transition rates so that they are general in this respect. In the case the growth velocity is positive, the affinity (Eq. **15**) can be interpreted as the entropy production per added monomer. The affinity has two contributions: the first from the driving force $\varepsilon$ given in terms of the mean free enthalpy of copolymerization and the second due to the disorder (Eq. **10**) of the monomers in the chain. This shows in particular that the copolymer can grow by an entropic effect of disorder even when the driving forces are slightly negative. Indeed, the randomness incorporated in the chain has a non-negative contribution in the form of the disorder (Eq. **10**) to the thermodynamic force or affinity per monomer (Eq. **15**), showing how the stored pattern can influence back the dynamics of the system. Equilibrium occurs when the affinity (Eq. **15**) vanishes, $A_{\mathrm{eq}} = 0$, so that the mean driving force is equal to minus the equilibrium disorder. In this case, no systematic growth is possible and each monomer appears with its equilibrium distribution.

In general, the transition rates $W(\omega, \omega')$ depend on the structure and composition of the whole chain $\omega$ so that nonlocal or cooperative effects can generate long-range correlations along the chain, thereby reducing the disorder (Eq. **10**). However, in many cases, the transition rates only depend on the few monomers at the end of the chain, in particular, if the copolymerization process is mainly controlled by the free enthalpy changes of bond formation in the reactions $\omega + m \rightleftharpoons \omega m$. In the most local process, the transition rates only depend on the last monomer $m$ which is added to or removed from the chain. In this case, the transition rates are denoted as $W(\omega, \omega m) = k_{+m}$ and $W(\omega m, \omega) = k_{-m}$ and the quantities

$$\varepsilon_m = \ln\frac{k_{+m}}{k_{-m}} \qquad \textbf{(16)}$$

represent the driving forces favoring the chain growth. They are given by the free enthalpy changes of the chemical reactions and are measured in units of the thermal energy: $\varepsilon_m = [G(\omega) - G(\omega m)]/(k_{\mathrm{B}}T)$. In this respect, the driving forces incorporate the chemical potentials of the monomers so that they can be varied by changing control parameters such as the chemical concentrations of the different monomers in solution around the copolymer chain.

Under the assumption of Eq. **16**, the mean value of the driving forces can be calculated as $\varepsilon = \sum_m \mu_1(m)\varepsilon_m$ by averaging them over the normalized distribution $\mu_1(m)$ of monomers. If we further assume that there is no free enthalpy difference between the chains $\omega m$, the driving forces are equal to each other, $\varepsilon_m = \varepsilon$ for all $m$, whereupon equilibrium occurs when the driving force takes the value $\varepsilon_{\mathrm{eq}} = -\ln M$ and the disorder reaches its maximum value $D_{\mathrm{eq}} = \ln M$. On the other hand, we notice that, for given positive values of the thermodynamic entropy production and the growth velocity, the disorder can be reduced at the expense of the driving force $\varepsilon$ according to Eq. **15**. Moreover, in the limit of an arbitrarily high driving force $\varepsilon$ no detachment of monomers occurs and the disorder reaches the value $D = -\sum_m \eta_m \ln \eta_m$, which is typically smaller than the equilibrium disorder $D_{\mathrm{eq}}$. Therefore, the nonequilibrium drive may contribute to some extend to the ordering of the copolymer.

We next consider a copolymerization process under the influence of a template, which determines the probabilities to add or remove monomers (see Fig. **1b**). This is typically the case during DNA transcription or replication processes where the DNA sequence influences the probabilities to add the monomers, favoring the Watson-Crick pairing rule A-T and C-G. The template is thus composed of a chain $\alpha = \alpha_1\alpha_2 \ldots \alpha_l \ldots$ and the transition rates now depend on this underlying substrate: $W(\omega, \omega'|\alpha)$. In this case, the previous considerations are modified in the following way. The entropy (Eq. **3**), the velocity (Eq. **6**), and the mean driving force are now averaged not only over the distribution of monomers but also over the template. We suppose that the template is described by a stationary distribution $\nu_l(\alpha) = \nu_l(\alpha_1\alpha_2\cdots\alpha_l)$ which is normalized to unity $\sum_\alpha \nu_l(\alpha) = 1$ for all the chains of length $l$. In the stationary regime, the probability to have a chain $\omega$ at time $t$ here takes the form

$$P_t(\omega) = p_t(l)\,\mu_l(\omega|\alpha) \qquad \textbf{(17)}$$

at long times, where $\mu_l(\omega|\alpha)$ is the distribution of chains $\omega$ of length $l$ grown on the template $\alpha$. We expect that the template ensemble average corresponds to the average over a typical realization of the template sequence. We can here also introduce the mean entropy (Eq. **9**) and enthalpy per monomer by averaging over the template. The growth speed (Eq. **6**) is now obtained with a long typical sequence of the template.

On the other hand, the disorder (Eq. **10**) now becomes the Shannon conditional entropy of the copolymer with respect to the template:

$$\begin{aligned} D(\text{polymer}|\,\text{template}) \\ = \lim_{l\to\infty} -\frac{1}{l} \sum_\alpha \nu_l(\alpha) \sum_\omega \mu_l(\omega|\alpha) \ln \mu_l(\omega|\alpha)\,. \end{aligned} \qquad \textbf{(18)}$$

The conditional entropy between two random variables can also be expressed in terms of the mutual information $I(\mathrm{X},\mathrm{Y}) = D(\mathrm{X}) - D(\mathrm{X}|\mathrm{Y})$ between the two variables [18]. The mutual information is bounded according to $0 \leq I(\mathrm{X},\mathrm{Y}) \leq \min\{D(\mathrm{X}), D(\mathrm{Y})\}$ and measures how much the knowledge on one of these variables reduces the uncertainty about the other. It is zero when the two variables are independent so that we recover the previous results (Eq. **10**) and (Eq. **15**) when the substrate does not influence the copolymerization process. Accordingly, the affinity here becomes

$$\begin{aligned} A &= \varepsilon + D(\text{polymer}|\,\text{template}) \\ &= \varepsilon + D(\text{polymer}) - I(\text{polymer}, \text{template}) \end{aligned} \qquad \textbf{(19)}$$

and is thus directly expressed in terms of the mutual information that the copolymer acquires from its template or more generally from its environment. The mean driving force is given by $\varepsilon = -g/T$ as in Eq. **15**.

To be specific, we can take transition rates of the form $W(\omega, \omega m | \alpha) = k_{+mn}$ and $W(\omega m, \omega | \alpha) = k_{-mn}$ if $n = \alpha_{l+1}$ is the template at the position $l + 1$ and $\omega$ is a chain of length $l$. The corresponding driving forces

$$\varepsilon_{mn} = \ln \frac{k_{+mn}}{k_{-mn}} \qquad \textbf{(20)}$$

now depend on the substrate as well. Accordingly, in the case of Eq. **20**, the mean driving force is given by $\varepsilon = \sum_{m,n} \nu_1(n) \mu_1(m|n) \varepsilon_{mn}$, where $\nu_1(n)$ is the distribution of monomers on the template.

If the copolymer grows and $v > 0$, the affinity is the entropy production per added monomer so that Eq. **19** shows that substantial information can be generated if the mean driving force $\varepsilon$ is large enough. We illustrate this fundamental result with DNA replication in the next section.
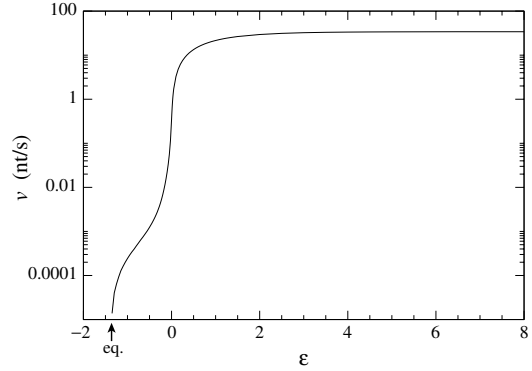
## DNA Replication

In this section, we consider the process of DNA replication. In this case, the subunits of the growing polymer and the template are the four nucleotides N = A, T, C, or G. The monomers are the corresponding nucleoside triphosphates NTP, which drive the reaction by their high free enthalpy relative to the linked nucleotides. It turns out that the DNA polymerase copies the DNA with a fairly small error rate. In the actual DNA replication, an exonuclease will act as a proofreading mechanism in order to correct possible errors. Proofreading allows one to decrease the error rate up to the discrimination squared [19, 20, 21]. This mechanism can be modeled by considering several reactions $\rho$ during the copolymerization, namely $W_\rho(\omega, \omega' | \alpha)$. These additional pathways will change the average driving force and speed, but the disorder (Eq. **18**) will remain unchanged for a given distribution of nucleotides.
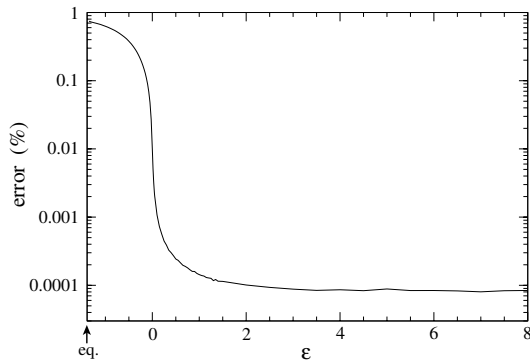
**Fig. 2.** Percentage of misincorporations as a function of the driving force $\varepsilon$. The minimal error rate is given by $1 - \nu_A \eta_{TA} - \nu_T \eta_{AT} - \nu_C \eta_{GC} - \nu_G \eta_{CG} \simeq 8.3 \, 10^{-5}$. The maximal error rate occurs at equilibrium where detailed balance conditions are satisfied, according to which each nucleotide insertion is balanced by its removal from the chain. Under this condition and if $\varepsilon_{mn} = \varepsilon$ for all $m$ and $n$, every nucleotide is included with equal probability and the error rate is $0.75$. The kinetic constants of Watson-Crick pairing are taken to be $k_{+TA} = 25 \, \text{s}^{-1}$, $k_{+AT} = 45 \, \text{s}^{-1}$, $k_{+GC} = 37 \, \text{s}^{-1}$, and $k_{+CG} = 43 \, \text{s}^{-1}$ [22]. The discrimination between the nucleotide $m$ and the template $n$ is defined as $d_{mn} = k_{+mn}/k_{+rn}$ where $r$ denotes the "correct" nucleotide forming the Watson-Crick pair $rn$ with the template nucleotide $n$. The discriminations take the values $d_{AA} = d_{AG} = d_{GA} = 1/280000$, $d_{CA} = 1/210000$, $d_{TT} = 1/250000$, $d_{CT} = 1/570000$, $d_{GT} = 1/3600$, $d_{AC} = 1/71000$, $d_{TC} = 1/640000$, $d_{CC} = 1/2300000$, $d_{TG} = 1/59000$, and $d_{GG} = 1/110000$. The reversed kinetic constants are taken as $k_{-mn} = k_{+mn} e^{-\varepsilon}$ according to Eq. **20** with $\varepsilon_{mn} = \varepsilon$ incorporating the monomer concentrations (see text section on dissipation-information tradeoff).

To analyze purely kinetic effects, we will assume in the following that no free enthalpy difference exists between correct and incorrect chains, implying $\varepsilon_{mn} = \varepsilon$. We will thus consider the different effects as a function of this parameter. The transition rates and the driving forces (Eq. **20**) incorporate external conditions such as the chemical concentrations of the nucleotides, the polymerase cofactors, and the products of the polymerization. Accordingly, the driving force $\varepsilon$ will typically be a function of such control parameters.

**Fig. 3.** Velocity of the replication process as a function of the driving force $\varepsilon$. The maximal speed is given by about $34$ nucleotides per second as explained in the text section on DNA replication.
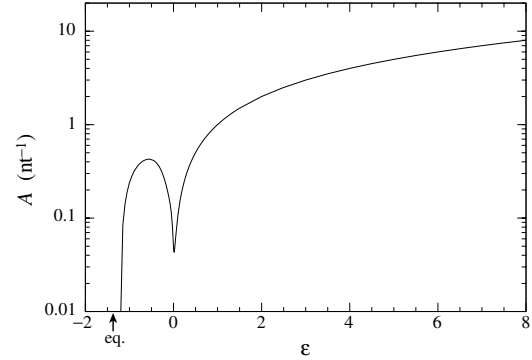
**Fig. 4.** Affinity per copied nucleotide as a function of the driving force $\varepsilon$. The local minimum is around $\varepsilon \simeq 0.015$ and indicates the transition between the error driven regime and the externally driven regime. Since the growth speed is positive, this affinity is the entropy production par copied nucleotide.
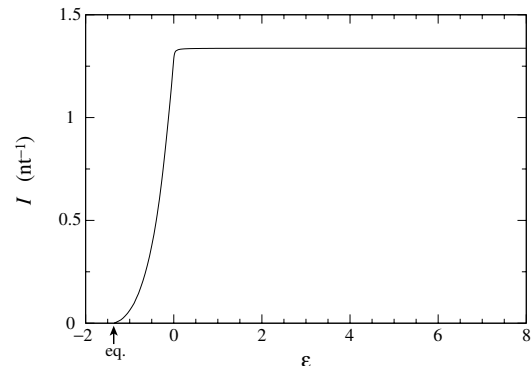
**Fig. 5.** Mutual information between the copied DNA strand and the original strand, as a function of the driving force $\varepsilon$. The saturation occurs after the transition between the error driven and the externally driven regimes. The value of the mutual information at saturation is $I_{\max} \simeq 1.337$ nats.

For concreteness, we consider the case of the DNA polymerase Pol $\gamma$, which replicates the human mitochondrial DNA. Forward kinetic constants for the incorporation of both correct and incorrect nucleotides are available [22] and we used these for our simulations. The human mitochondrial DNA is 16.5 kb long and can be obtained from GenBank [23]. To have a good statistical estimation of the disorder (Eq. **18**), we used longer DNA sequences generated from the same triplet distribution as the original mitochondrial DNA. Random trajectories corresponding to the master equation (Eq. **1**) are obtained with Gillespie's algorithm [24].

Since we assumed no free enthalpy difference between the chains, equilibrium occurs when $\varepsilon = -\ln 4$, where each nucleotide is inserted with equal probability regardless of the underlying template. Therefore, at equilibrium, the velocity and the mutual information vanish, while the error rate is 0.75.

In Fig. 2, we depict the fraction of errors as a function of the driving force and, in Fig. 3, the velocity of the replication process. The key result is that the fraction of errors is a decreasing function of the driving force. The minimal error rate one can achieve is given in terms of the probability $\eta_{mn} = k_{+mn}/\sum_m k_{+mn}$ for a base $m$ on a template $n$ and occurs in the infinite dissipation limit, i.e., the limit $\varepsilon \to \infty$. Nucleotides are then inserted with rates $k_{+mn}$ but never detach, as shown by Eq. **20**. Figure 2 illustrates the important point that polymerases must operate far from equilibrium in order to achieve a high fidelity in the replication of DNA information; the same observation is expected for various biological entities that must present coherent space-time behaviors such as molecular motors. The velocity increases with the driving force, reaching a maximal speed given by $v_{\max}^{-1} = \sum_n \nu_1(n)/(\sum_m k_{+mn})$ which is here of about 34 bases per second. Also, both the velocity and the error rate undergo a steep transition as the driving force $\varepsilon$ becomes positive.

In Fig. 4, we represent the affinity (Eq. **15**) per copied nucleotide, which gives the corresponding entropy production since the growth velocity is here positive. This curve is non monotonic and presents a minima around $\varepsilon \simeq 0.015$. On the left-hand side of it, the external driving force is weak or negative and dissipation occurs chiefly due to the positive driving caused by the incorporation of errors whereas, on the right-hand side, dissipation is mainly due to the external driving force. This transition corresponds to the sharp behavior observed in the error rate and in the velocity of the process. The same behavior is also observed in the mutual information $I$ between the original and the copied DNA strands, as depicted in Fig. 5. It increases with the driving force, nearly reaching its maximal value $I_{\max} = \sum_{m,n} \eta_{mn}\nu_1(n)\ln[\eta_{nm}/\sum_n \eta_{mn}\nu_1(n)] \simeq 1.337$ nats after the threshold point. Afterwards, nearly all the original information has been acquired. The result is that the increase of mutual information corresponds to a reduction of disorder and allows fidelity in the transfer of information. Contrary to the situation at equilibrium where this transfer is not possible, the nonequilibrium conditions turn out to induce the generation of mutual information.

## Conclusions

In this paper, we have developed the nonequilibrium statistical thermodynamics of copolymerization processes, delineating the role of configurational disorder in copolymer chains. This latter plays the role of a positive driving force favoring copolymer's growth. In particular, disorder arises from the intricate interplay between fluctuations and nonequilibrium conditions, reaching its minimal value in the infinite dissipation limit. Furthermore, we have shown how information can be generated out of equilibrium in these fluctuating processes. In this regard, we have observed a transition between two regimes. At small or negative external driving forces, the growth process is essentially driven by the incorporation of errors in the copolymer chain while, beyond a threshold, it is driven by free enthalpy sources. After this transition point, it turns out that the extra dissipation does not lead to a substantial increase in the fidelity of the replicated strands, as measured by the error rate or the mutual information between the original and the copied DNA strand. Information acquisition thus appears to have an energy cost analogous to information erasure [25, 26, 27]. Similar or enhanced effects are expected in the presence of autocatalysis [10], possibly, with different dependences on control parameters.

Albeit the results apply to every copolymerization process, they are fundamental for DNA replication and transcription and the understanding of information acquisition by biological entities. Biological systems are characterized by several properties such as metabolism and self-reproduction. Each of them can be said to be encoded in the genetic code. While metabolism refers to nonequilibrium thermodynamics, self-reproduction involves information processing. The present work establishes an essential connection between both properties on the basis of nonequilibrium statistical thermodynamics. In this perspective, dissipation allows the emergence of temporal ordering and the spontaneous generation of information, explaining in this way a key feature of complex systems [28]. Information generation can thus be understood as resulting from the reduction of disorder in copolymers growing out of equilibrium and the ability to store and retrieve information in nonequilibrium environments. If furthermore this information can be copied and processed back by the nonequilibrium dynamics of the system, the self-reproduction of biological systems could be explained. From this viewpoint, we could say that biological systems are physico-chemical systems with a built-in thermodynamic arrow of time.

1. Mills DR , Peterson RL, and Spiegelman S (1967) An extracellullar Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl. Acad. Sci. USA* 58:217-224.

2. Biebricher CK and Gardiner WC (1997) Molecular Evolution of RNA in vitro. *Biophysical Chemistry* 66:179-192.

3. Eigen M and Schuster P (1979) *The hypercycle - a principle of natural self-organization* (Springer, Heidelberg).

4. Eigen M (1992) *Steps towards Life: A Perspective on Evolution* (Oxford University Press, Oxford).

5. Gaspard P (2004) Time-reversed dynamical entropy and irreversibility in Markovian random processes. *J. Stat. Phys.* 117:599-615.

6. Gaspard P (2007) Temporal ordering of nonequilibrium fluctuations as a corollary of the second law of thermodynamics. *C. R. Physique* 8:598-608.

7. Andrieux D, Gaspard P, Ciliberto S, Garnier N, Joubaud S, and Petrosyan A (2007) Entropy production and time asymmetry in nonequilibrium fluctuations. *Phys. Rev. Lett.* 98:150601.

8. Andrieux D, Gaspard P, Ciliberto S, Garnier N, Joubaud S, and Petrosyan A (2008) Thermodynamic time asymmetry in non-equilibrium fluctuations. *J. Stat. Mech.* P01002.

9. Nicolis G, Subba Rao G, Subba Rao J, and Nicolis C (1989) Generation of spatially asymmetric, information-rich structures in far from equilibrium systems. In *Structure, coherence and chaos in dynamical systems*, eds Christiansen PL, Parmentier RD (Manchester University Press, Manchester and New York) pp. 287-299.

10. Goldbeter A and Nicolis G (1972) Far from Equilibrium Synthesis of Small Polymer Chains and Chemical Evolution. *Biophysik* 8: 212-226.

11. Schnakenberg J (1976) Network theory of microscopic and macroscopic behavior of master equation systems. *Rev. Mod. Phys.* 48:571-585.

12. Nicolis G and Prigogine I (1977) *Self-Organization in Nonequilibrium Systems* (Wiley, New York).

13. Hill TL (2005) *Free Energy Transduction and Biochemical Cycle Kinetics* (Dover, New York).

14. Gaspard P (2004) Fluctuation theorem for nonequilibrium reactions. *J. Chem. Phys.* 120:8898-8905.

15. Luo J-L, Van den Broeck C, and Nicolis G (1984) Stability criteria and fluctuations around nonequilibrium states. *Z. Phys. B - Condensed Matter* 56:165-170.

16. Coleman BD and Fox TG (1963) General Theory of Stationary Random Sequences with Applications to the Tacticity of Polymers. *J. Polym. Sci. A* 1: 3183-3197.

17. Coleman BD and Fox TG (1963) Multistate Mechanism for Homogeneous Ionic Polymerization. I. The Diastereosequence Distribution. *J. Chem. Phys.* 38: 1065-1075.

18. Chaitin GJ (1979) Toward a mathematical definition of "life". In *The Maximum Entropy Formalism*, eds Levine RD, Tribus M (MIT Press, Cambridge) pp. 477-498.

19. Hopfield JJ (1974) Kinetic Proofreading: A New Mechanism for Reducing Errors in Biosynthetic Processes Requiring High Specificity. *Proc. Natl. Acad. Sci. USA* 71:4135-4139.

20. Ninio J (1975) Kinetic amplification of enzyme discrimination. *Biochimie* 57:587-595.

21. Bennett CH (1979) Dissipation-error tradeoff in proofreading. *Biosystems* 11:85-91.

22. Lee H and Johnson K (2006) Fidelity of the Human Mitochondrial DNA Polymerase. *J. Biol. Chem.* 281:36236-36240.

23. Homo sapiens DNA mitochondrion, 16569 bp, locus AC 000021, version GI:115315570, `http://www.ncbi.nlm.nih.gov`.

24. Gillespie DT (1976) A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *J. Comp. Phys.* 22:403-434.

25. Landauer R (1961) Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* 5:183-191.

26. Bennett CH (1982) The Thermodynamics of Computation - A Review. *Int. J. Theor. Phys.* 21:905-940.

27. Andrieux D and Gaspard P (2008) Dynamical randomness, information, and Landauer's principle. *Europhys. Lett.* 81:28004.

28. Nicolis G and Nicolis C (2007) *Foundations of Complex Systems* (World Scientific, Singapore).