Nonequilibrium Phase Transitions Associated with DNA Replication

Hyung-June Woo* and Anders Wallqvist

Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Fort Detrick, Maryland 21702, USA

(Received 26 August 2010; published 9 February 2011)

Thermodynamics governing the synthesis of DNA and RNA strands under a template is considered analytically and applied to the population dynamics of competing replicators. We find a nonequilibrium phase transition for high values of polymerase fidelity in a single replicator, where the two phases correspond to stationary states with higher elongation velocity and lower error rate than the other. At the critical point, the susceptibility linking velocity to thermodynamic force diverges. The overall behavior closely resembles the liquid-vapor phase transition in equilibrium. For a population of self-replicating macromolecules, Eigen's error catastrophe transition precedes this thermodynamic phase transition during starvation. For a given thermodynamic force, the fitness of replicators increases with increasing polymerase fidelity above a threshold.

DOI: 10.1103/PhysRevLett.106.060601 PACS numbers: 05.70.Ln, 87.10.Mn, 87.14.gk

The replication of genes by biological organisms lies at the core of their activities, seemingly driving competitions for survival [1]. The process on the molecular level is driven by enzymes (polymerases) catalyzing the growth of a DNA primer strand (the nascent chain of nucleotides complementary to the template strand) based on the Watson-Crick base pairing rules (A is paired with T and G with C). Because of their movement along the chain, polymerases belong to molecular motors converting chemical free energy into work [2,3]. The copying process, however, involves errors, whose rate varies with the replication conditions as has been observed recently [4–7]. This thermodynamic interpretation of replication may be relevant to many biological phenomena: mutation rates of bacteria increase under stress such as starvation [8]. RNA viruses maintain their mutation rates near the threshold of error catastrophe, beyond which populations cannot sustain stable genomes [9,10].

Compared to a polymerase-nucleotide strand complex (''replicator'') far from equilibrium primarily driven by external chemical potentials, a replicator near equilibrium copies its genome with higher error rates, which make extra contributions to the entropy production [4]. In this Letter, we show that the transition between these externally driven and disorder-driven regimes of replication exhibits all characteristic features of phase transitions one normally expects in equilibrium. Many systems driven out of equilibrium exhibit phase transitionlike behavior, including models with absorbing states [11] and the evolution of opinions in social networks [12]. The features observed here, however, are unique in the sense that they offer a close nonequilibrium analog of the liquid-vapor transition. We combine this single molecule thermodynamics with the population dynamics of replicators [9], which provides a thermodynamic interpretation of molecular evolution.

Considering first macroscopic thermodynamics of a single replicator in a reservoir, the chemical reaction involved is

$$(\mathrm{DNA})_N + \mathrm{M} \rightleftharpoons (\mathrm{DNA})_{N+1} + \mathrm{P}, \tag{1}$$

where $(\mathrm{DNA})_N$ is the primer strand of length N, M denotes one of four nucleoside triphosphate (NTP) monomers, and P is pyrophosphate (PPi). The entropy of the system plus surrounding $S = S(N_m, N_p, N)$ is a function of the number of free monomers N_m, the number of PPi N_p, and the chain length N in units of monomer size. Their changes are related by $dN_m = -dN_p = -dN$. The entropy change (or production) can be written as [13]

$$dS = -\frac{\mu_m}{T} dN_m - \frac{\mu_p}{T} dN_p + \frac{f}{T} dN = \frac{f - \Delta G}{T} dN, \tag{2}$$

where $\mu_i = -T\partial S/\partial N_i$ are the chemical potentials, T is temperature, $f = T\partial S/\partial N$ is the external force acting on the primer strand, and $\Delta G = \mu_p - \mu_m = \Delta G^\circ - T \ln([\mathrm{M}]/[\mathrm{P}])$ is the Gibbs energy change of reaction (1) (Boltzmann constant is 1). The entropy production rate $\dot{S}$ can thus be written as

$$\dot{S} = F v, \tag{3}$$

where $v = dN/dt$ is the elongation velocity and the thermodynamic force F imposed by the reservoir is given by

$$F = \frac{f - \Delta G}{T}. \tag{4}$$

The two terms in Eq. (4) represent mechanical and chemical driving forces of the reservoir. We have assumed that the reaction and movements are tightly coupled: Eqs. (3) and (4) correspond to a special case of a more general expression for molecular motors [Eq. (7) of Ref. [2] with reaction rate equal to v]. Forces acting on nucleotide strands or polymerase can be controlled with

optical tweezers [14], and we assumed that the direction of f coincides with that of elongation.

Stochastic dynamics allows us to relate this thermodynamic description to molecular features [2,4,5] and derive the constitutive relation between F and v. The attachment (detachment) rates of monomer α base-paired to β are denoted $k_{\pm}(\alpha|\beta)$, where α and β are any of m possible monomer types [$m = 4$ for DNA (RNA)]. The conditional probability $P_n(\alpha^n, t|\beta)$ of finding the chain of length n at time t with sequence $\alpha^n \equiv (\alpha_n, \alpha_{n-1}, \ldots, \alpha_1)$ under a given template sequence β satisfies

$$\begin{aligned}\frac{d}{dt}P_n(\alpha^n, t|\beta) &= k_+(\alpha_n|\beta_n)P_{n-1}(\alpha^{n-1}, t|\beta) \\ &+ \sum_{\alpha'} k_-(\alpha'|\beta_{n+1})P_{n+1}(\alpha'\alpha^n, t|\beta) \\ &- k_-(\alpha_n|\beta_n)P_n(\alpha^n, t|\beta) \\ &- \sum_{\alpha'} k_+(\alpha'|\beta_{n+1})P_n(\alpha^n, t|\beta). \end{aligned} \tag{5}$$

We write $P_n(\alpha^n, t|\beta) = p_n(t)q_n(\alpha^n|\beta)$ [4], where $p_n(t)$ is the probability of having chain length n at time t. Since the rate constants $k_{\pm}(\alpha|\beta)$ in Eq. (5) are all local functions and do not depend on neighboring nucleotides, we expect

$$q_n(\alpha^n|\beta) = q(\alpha_n|\beta_n)q(\alpha_{n-1}|\beta_{n-1})\cdots q(\alpha_1|\beta_1) \tag{6}$$

in the stationary state, where $q \equiv q_1$. Using Eq. (6) in Eq. (5), assuming that the template sequence distribution is spatially uncorrelated, and averaging both sides over β_{n+1}, we can write

$$\begin{aligned}\frac{dp_n}{dt}q(\alpha|\beta) &= k_+(\alpha|\beta)p_{n-1} + J_-q(\alpha|\beta)p_{n+1} \\ &- k_-(\alpha|\beta)q(\alpha|\beta)p_n - J_+q(\alpha|\beta)p_n, \end{aligned} \tag{7}$$

where $J_+ = \sum_\alpha \langle k_+(\alpha|\beta)\rangle_\beta$, $J_- = \sum_\alpha \langle k_-(\alpha|\beta)q(\alpha|\beta)\rangle_\beta$ are the rates of chain growth and shrinkage.

Summing both sides of Eq. (7) over n,

$$k_+(\alpha|\beta) + [J_- - k_-(\alpha|\beta) - J_+]q(\alpha|\beta) = 0, \tag{8}$$

whereas multiplying both sides of Eq. (7) by n and summing over n using $v = \sum_n n dp_n/dt$, we have

$$q(\alpha|\beta)v = k_+(\alpha|\beta) - J_-q(\alpha|\beta). \tag{9}$$

If we sum Eq. (9) over α and average over β, we get

$$v = J_+ - J_-, \tag{10}$$

where $\sum_\alpha q(\alpha|\beta) = 1$ has been used. Equations (8) and (10) give

$$q(\alpha|\beta) = \frac{k_+(\alpha|\beta)}{k_-(\alpha|\beta) + v}. \tag{11}$$

A substitution of Eq. (11) into the normalization condition $\sum_\alpha \langle q(\alpha|\beta)\rangle_\beta = 1$ yields

$$\sum_\alpha \left\langle \frac{k_+(\alpha|\beta)}{k_-(\alpha|\beta) + v}\right\rangle_\beta = 1, \tag{12}$$

which determines the velocity implicitly. From Eq. (11), the fraction ϵ (error rate) of monomers for which $\alpha \neq \beta^\dagger$, where $\beta^\dagger$ is the correct Watson-Crick complementary base of β, can be obtained by

$$\epsilon = \left\langle \sum_{\alpha\neq\beta^\dagger} q(\alpha|\beta)\right\rangle_\beta. \tag{13}$$

The backward rate k_- can be written in terms of the forward rate k_+ introducing $g_{\alpha\beta}$ by

$$k_-(\alpha|\beta) = e^{-g_{\alpha\beta}/T}k_+(\alpha|\beta). \tag{14}$$

The parameter $-g_{\alpha\beta}$ corresponds to the binding free energy of a single monomer α against the template base β. This free energy, however, is a potential of mean force for the given fixed nucleotide pair. Because of the contribution of sequence disorder (see below), this parameter is in general not equal to the change in external chemical potentials ΔG. The entropy production per monomer added is [4]

$$F = \left\langle \sum_\alpha \left[\frac{g_{\alpha\beta}q(\alpha|\beta)}{T} - q(\alpha|\beta)\ln q(\alpha|\beta)\right]\right\rangle_\beta \equiv \frac{\langle g_{\alpha\beta}\rangle}{T} + D, \tag{15}$$

the first term arising from the consumption of monomers and $D = -\sum_\alpha \langle q(\alpha|\beta)\ln q(\alpha|\beta)\rangle_\beta$ from the sequence disorder, the Gibbs entropy of the uncertainty of α paired to a known β due to copying errors. Comparing Eqs. (4) and (15), we note that if $D = 0$, $\langle g_{\alpha\beta}\rangle = f - \Delta G$: the average of $g_{\alpha\beta}$ coincides with the external thermodynamic driving force. With $D \neq 0$, however, the experimentally controllable parameter is the thermodynamic force F given by Eq. (4) rather than $\langle g_{\alpha\beta}\rangle$. Combining Eqs. (12) and (14), we observe that $v = 0$ when $\sum_\alpha \langle e^{g_{\alpha\beta}/T}\rangle_\beta = 1$, whereas if $g_{\alpha\beta} \to \infty$, $v \to J_+$.

To derive explicit formulas for stationary properties, we adopt the following simple model:

$$\begin{aligned} k_+(\alpha|\beta) &= \delta_{\alpha\beta^\dagger}k + (1 - \delta_{\alpha\beta^\dagger})\gamma k, \\ g_{\alpha\beta}/T &= g - (1 - \delta_{\alpha\beta^\dagger})\ln\eta, \end{aligned} \tag{16}$$

and a uniform distribution of β, where $0 < \gamma < 1$ and $\eta \geq 1$ are the ratios of monomer insertion rates k_+ and dissociation constants k_-/k_+ for incorrect to correct base pairs, respectively, (the magnitude of binding free energy is smaller for incorrect pairs). The equilibrium occurs when $g = g_{\rm eq} \equiv -\ln[1 + (m-1)/\eta]$. Equation (12) gives for $\bar{v} = v/k$,

$$\bar{v} = \Gamma + \sqrt{\Gamma^2 + \gamma e^{-g}(m - 1 + \eta - e^{-g}\eta)}, \quad (17a)$$

$$\Gamma = \tfrac{1}{2}[1 - e^{-g} - \gamma + \gamma(m - e^{-g}\eta)]. \quad (17b)$$

The error rate from Eq. (13) in this case becomes

$$\epsilon = \frac{m - 1}{e^{-g}\eta + \bar{v}/\gamma}, \quad (18)$$

while Eq. (15) gives

$$F = \frac{\ln(1 + e^{g}\bar{v})}{e^{-g} + \bar{v}} + (m - 1)\frac{\ln(1 + e^{g}\bar{v}/\eta\gamma)}{e^{-g}\eta + \bar{v}/\gamma}. \quad (19)$$

It is useful to consider the limit of perfect selectivity $\gamma \to 0$. The values of $\bar{v}$, ϵ, and F in this limit are $\bar{v}_0 = 1 - e^{-g}$ if $g > 0$ and $\bar{v} = 0$ otherwise,

$$\epsilon_0 = \begin{cases} 1 - e^{g} & \text{if } g_{\text{eq}} \le g \le 0, \\ 0 & \text{if } g \ge 0, \end{cases} \quad (20)$$

$$F_0 = \begin{cases} (1 - e^{g})\ln\left[\left(\frac{m-1}{e^{-g}-1}\right)\frac{1}{\eta}\right] & \text{if } g \le 0, \\ g & \text{if } g \ge 0. \end{cases} \quad (21)$$

Figure 1 shows the dependence of stationary properties on g. The singular behavior of F versus g was observed from numerical simulations in Ref. [4], which is revealed here analytically. Near equilibrium, a polymerase can attach (detach) monomers with two different relative rates for a given F [Fig. 1(c)]. We also tested Eq. (6) by direct Monte Carlo simulations of Eq. (5), where chains were grown stochastically under a random template sequence β with the initial condition $n = 0$. The numerical results shown in Fig. 1 confirm Eq. (6).

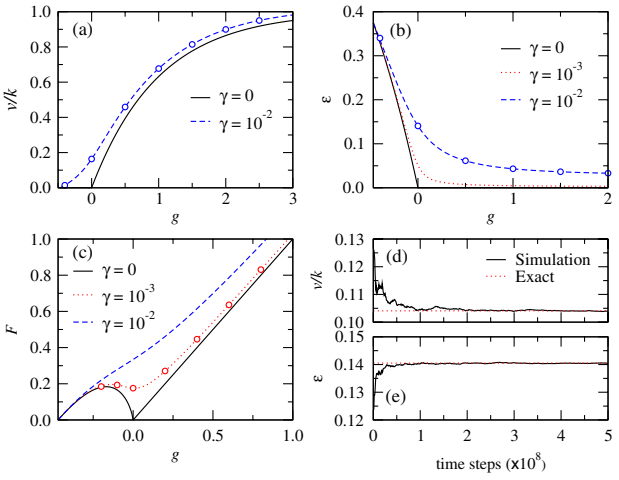

FIG. 1 (color online). Dependence of stationary properties on g. (a) $\bar{v}$ for $m = 4$ and $\eta = 5$; (b) ϵ for $m = 4$ and $\eta = 5$; (c) F for $m = 4$ and $\eta = 5$; (d) convergence of $\bar{v}$ in numerical simulations for $m = 2$, $\eta = 10$, $\gamma = 10^{-3}$, and $g = 0.1$; and (e) convergence of ϵ in simulations for $m = 4$, $\eta = 5$, $\gamma = 10^{-2}$, and $g = 0$. Size of time step in Monte Carlo was 0.01 in units of k^{-1}. Circles in (a)–(c) are from the numerical simulations.

Equation (19) should match Eq. (4) in stationary states, and Eqs. (17)–(19) give the dependence of v and ϵ on F parametrically through g [Fig. 2], which resembles the p-V isotherms of van der Waals gas: below a critical γ_c, the system exhibits multiple values of v and ϵ for a given F. Most polymerases have low mismatch rates ($\gamma \lesssim 10^{-4}$) and are expected to be subcritical. As $\gamma \to 0$, Eqs. (20) and (21) give

$$F = \epsilon_0 \ln[(m - 1)(\epsilon_0^{-1} - 1)/\eta]. \quad (22)$$

As in equilibrium thermodynamics, it is reasonable to rule out the segment of v and ϵ values violating the stability criterion for the susceptibility $\chi = \partial v/\partial F > 0$. When F is lowered (e.g., from starvation), one reaches a spinodal where $\chi \to \infty$, and v and ϵ discontinuously jump to smaller and larger values, respectively. It is expected that the distributions of v and ϵ would become unusually broad at the critical point where $\partial F/\partial v = \partial^2 F/\partial v^2 = 0$. Figure 3 shows the phase diagram analogous to the p-T diagram of fluids, where the stability limit of the low-error phase terminates at the critical point (F_c, γ_c), which shifts to smaller values with increasing η.

How will this behavior of a single replicator affect the population dynamics of competing molecules? Such a population exists as a quasispecies, a cloud of variable genotypes surrounding a master sequence [9,15–17], whose evolution is described by $dn_i/dt = \sum_j Q_{ij} r_j n_j$, where n_i is the number of replicators with genotype i of length L, $r_i = \bar{v}k_i/L$ is its replication rate ("fitness") with NTP-incorporation rate k_i, and $Q_{ij} = \epsilon^{d_{ij}}(1 - \epsilon)^{L - d_{ij}}$ is the mutation matrix between i and j with Hamming distance d_{ij} (the number of nucleotides that are different). The more common form of the quasispecies dynamics can be derived from $x_i = n_i/N$, where $N = \sum_i n_i$ is the total population size. Using $\sum_i Q_{ij} = 1$, one can also write $dN/dt = rN$, where $r = \sum_i r_i x_i$ is the mean fitness of the population. Equation (3) is generalized as

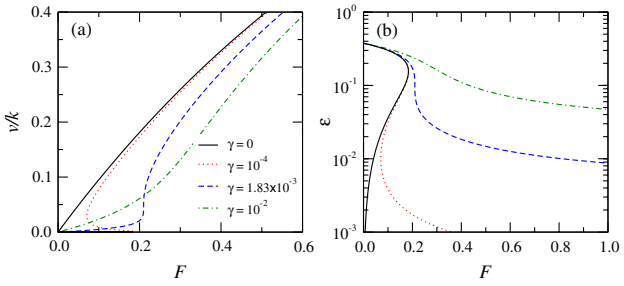

FIG. 2 (color online). (a) Velocity and (b) error rate as functions of F for $m = 4$ and $\eta = 5$. The solid line in (b) corresponds to Eq. (22). The critical point is reached when $\gamma_c = 1.83 \times 10^{-3}$.

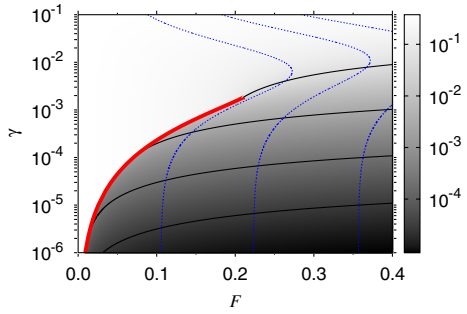

FIG. 3 (color online). Phase diagram for $m = 4$ and $\eta = 5$. Thick (red) line denotes the spinodal of high-v/low-ϵ phase terminating at the critical point $(F_c, \gamma_c) = (0.210, 1.83 \times 10^{-3})$. The gray scale indicates the error rate ϵ. Thin solid (black) lines show constant-ϵ contours at $\epsilon = 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}$. Dotted (blue) lines show constant-$\bar{v}$ contours at $\bar{v} = 0.1, 0.2, 0.3$ from left to right.

$$\dot{S} = F\sum_i n_i r_i L = NFLr, \tag{23}$$

which shows that fitness increases are driven by the entropy production.

The constitutive relation $r = r(F)$ on the population level can now be obtained from the solution of Eigen model in the infinite population limit [16]. For the simplest landscape $k_i/k = \delta_{i1}(A - 1) + 1$, where $i = 1$ denotes the master sequence with relative fitness $A > 1$, the mean fitness becomes

$$r = \frac{v}{L} \times \begin{cases} Ae^{-L\epsilon} & \text{if } \epsilon < \epsilon^*, \\ 1 & \text{otherwise,} \end{cases} \tag{24}$$

where

$$\epsilon^* = \frac{\ln A}{L} \tag{25}$$

is the error threshold [9]. For $\epsilon > \epsilon^*$, therefore, the constitutive relation on the population level remains the same and no nontrivial collective effect occurs. With ϵ below the threshold, on the other hand, $Lr > v$ and the population enhances its collective growth rate by maintaining the quasispecies peaked near the master sequence. Equation (25) can be combined with Eq. (18) to give boundaries of genomic stability in the phase diagram [Fig. 3]: $\epsilon(F, \gamma) < \ln A/L$. If F decreases during starvation under a constant $\gamma_t < \gamma < \gamma_c$, where γ_t corresponds to a "triple point," the error catastrophe transition ("melting") will be followed by the thermodynamic transition ("evaporation").

Can a population with highly error-prone polymerases, such as rudimentary ribozymes in an RNA environment [18], spontaneously evolve increasingly lower γ values and eventually enter the regime of genomic stability? Evolutionary walks with positive changes in r are thermodynamically favorable for a given F [Eq. (23)], which occurs in Fig. 3 for $\gamma \lesssim 10^{-2}$ where $\partial\bar{v}/\partial\gamma < 0$. An emergence of biological information from a population of random sequences may have required crossing the thermodynamic threshold where $\partial\bar{v}/\partial\gamma$ changes sign.

We assumed that the template sequence is spatially uncorrelated and mutations occur only via substitutions. Real genomic sequences show long range correlations often described by power laws, which can be modeled by considering insertion, deletion, and duplication events [19–21]. It will be of interest to extend the current study to examine such effects.

Funding support came from the Department of Defense High Performance Computing (HPC) Modernization Program Office under the HPC Software Applications Institute Initiative, the U.S. Army Medical Research and Materiel Command.

*woo@bioanalysis.org

[1] R. Dawkins, *The Selfish Gene* (Oxford, New York, 1989).
[2] F. Jülicher, A. Ajdari, and J. Prost, Rev. Mod. Phys. **69**, 1269 (1997).
[3] K. Kruse and K. Sekimoto, Phys. Rev. E **66**, 031904 (2002).
[4] D. Andrieux and P. Gaspard, Proc. Natl. Acad. Sci. U.S.A. **105**, 9516 (2008).
[5] D. Andrieux and P. Gaspard, J. Chem. Phys. **130**, 014901 (2009).
[6] C. Jarzynski, Proc. Natl. Acad. Sci. U.S.A. **105**, 9451 (2008).
[7] M. Esposito, K. Lindenberg, and C. Van den Broeck, J. Stat. Mech. (2010) P01008.
[8] I. Bjedov, O. Tenaillon, B. Gérard, V. Souza, E. Denamur, M. Radman, F. Taddei, and I. Matic, Science **300**, 1404 (2003).
[9] M. Eigen, Naturwissenschaften **58**, 465 (1971).
[10] J. P. Anderson, R. Daifuku, and L. A. Loeb, Annu. Rev. Microbiol. **58**, 183 (2004).
[11] H. Hinrichsen, Adv. Phys. **49**, 815 (2000).
[12] P. Holme and M. E. J. Newman, Phys. Rev. E **74**, 056108 (2006).
[13] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics* (Wiley, New York, 1985), 2nd ed.
[14] S. K. Perumal, H. Yue, Z. Hu, M. M. Spiering, and S. J. Benkovic, Biochim. Biophys. Acta **1804**, 1094 (2010).
[15] J. Swetina and P. Schuster, Biophys. Chem. **16**, 329 (1982).
[16] D. B. Saakian and C.-K. Hu, Proc. Natl. Acad. Sci. U.S.A. **103**, 4935 (2006).
[17] J.-M. Park and M. W. Deem, Phys. Rev. Lett. **98**, 058101 (2007).
[18] À. Kun, M. Santos, and E. Szathmáry, Nat. Genet. **37**, 1008 (2005).
[19] W. Li, Phys. Rev. A **43**, 5240 (1991).
[20] P. W. Messer, P. F. Arndt, and M. Lässig, Phys. Rev. Lett. **94**, 138103 (2005).
[21] D. B. Saakian, Phys. Rev. E **78**, 061920 (2008).