The thermodynamics of writing a random polymer

Christopher Jarzynski*

Department of Chemistry and Biochemistry, and Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742

The notion that information has physical, and in particular, thermodynamic, content can be traced to the paradox of Maxwell's demon, a sly creature who observes the microscopic motions of gas particles on both sides of a partition (1). By controlling a trap door the demon segregates fast particles from slow ones to create a temperature difference across the partition, seemingly without expending any work. Generations of physicists have scratched their heads over this apparent violation of the second law of thermodynamics (2–5). The resolution that has eventually emerged acknowledges that a real-life Maxwell's demon—say, a nanoscale machine designed for the task—collects information as it operates, and work must be expended to erase this information, otherwise the demon's memory banks fill up. The minimum work required is k_BT ln 2 per bit of information, precisely what is needed to rescue the second law from the paradox. In this issue of PNAS, Andrieux and Gaspard (6) analyze the flip side of the thermodynamic cost of information erasure; namely, the cost of information acquisition. The setting of their analysis is not a demon and a gas, but rather a process essential to living organisms: copolymerization, in which a chain-like molecule grows by the addition of chemically distinct units (monomers). The most celebrated example is the replication of DNA, by which genetic information is copied at the molecular level, ultimately to pass down the generations of a family tree.

Noting that copolymerization is a physical process "ruled by the statistical laws of motion and thermodynamics," Andrieux and Gaspard (6) set out to investigate the implications of these laws, focusing on the interplay between the information that gets stored in the sequence of monomers (e.g., the pattern of nucleotides *A*, *G*, *C*, and *T* in the case of DNA) and the thermodynamic forces that drive the copolymerization process. As a warm-up problem, they first consider polymer growth without a template. Imagine a solution containing an alphabet soup of monomer species, in which floats a single chain of length l, with chemical bonds linking adjacent units. This polymer can grow by a reaction that attaches a new monomer to its end or shrink by the reverse reaction, in which a monomer drops back into the solution. We further imagine that the concentrations of the free monomer species in solution are kept fixed by an external agent that pumps them into or out of solution as necessary. In this situation, the polymer might reach a nonequilibrium stationary state in which it grows at an average rate $v = \langle dl/dt \rangle > 0$,

Copolymerization is a physical process "ruled by the statistical laws of motion and thermodynamics."

feeding off the ever-replenished supply of monomers. Alternatively, the polymer might shrink ($v < 0$), losing monomers that are then pumped out of the solution, until it vanishes entirely. At the boundary between these two scenarios of gorging and starving is a third possibility: The monomeric concentrations might be just right to produce a state of equilibrium, with no tendency either to grow or to shrink ($v = 0$). What determines which behavior will in fact be observed?

Let us focus on the first of these alternatives: steady polymer growth. At a given instant in time, let ω denote the current polymer sequence (e.g., *ACC*···*GT*) and let $G(\omega)$ be the free enthalpy, or Gibbs free energy, of this polymer. As the sequence ω evolves in time, so does the free enthalpy, $G(\omega)$. Following ref. 6, let g denote the average change of free enthalpy per added monomer, in the stationary state of the growing polymer. Because—as we all learn in general chemistry—a reaction spontaneously proceeds downhill in free enthalpy, it is useful to think of the quantity

$$\varepsilon = -g/T \qquad [1]$$

as a thermodynamic force that drives polymer growth. We might then conclude that the state of steady polymer growth is characterized by a negative free enthalpy per monomer ($g < 0$), equivalently a positive force ε.

Not so fast, say Andrieux and Gaspard (6). The growing polymer records information, in the sequence of monomers, $\omega(t)$. There is randomness in this record, and this randomness must be included in the ledgers when making a proper thermodynamic accounting of the stationary state. More precisely, they argue that the condition for a growing polymer is not $\varepsilon > 0$, but rather

$$\varepsilon + D(\text{polymer}) > 0, \qquad [2]$$

where D(polymer) is the disorder, the average increase of Shannon entropy per monomer. [The Shannon entropy, a concept from coding theory, measures the degree of randomness of a string of symbols (7).] Thus, even when thermodynamic considerations seem to rule it out ($\varepsilon < 0$), polymer growth might be rescued by the "entropic effect of disorder" in the sequence, provided $\varepsilon + D > 0$!

To convince oneself that this conclusion is sensible, it is useful to grab a pencil and work through a simple, back-of-the-envelope example, in which this entropic effect compensates for a positive free enthalpy per monomer.

First, imagine a polymer in a dilute solution of a single monomer species, *A*. Let $k_{+A} = 0.9\ s^{-1}$ be the average rate at which new monomers join the chain, and let $k_{-A} = 1.0\ s^{-1}$ be the probability rate that a monomer at the end of the polymer will detach and fall back into the solution. Because detachments occur more frequently than attachments, a polymer *AAA*···*A* will steadily dissolve under these conditions. This is just what we would expect from an elementary thermodynamic analysis: The cost in free enthalpy to attach a single monomer is $g = g_A = -T \ln(k_{+A}/k_{-A}) \approx 0.1$ T; thus, polymer growth is "uphill" in free enthalpy, and the thermodynamic force is negative: $\varepsilon \approx -0.1$. Now, imagine that the solution also contains another species, *C*, with the same rate constants: $k_{+C} = 0.9\ s^{-1}$ and $k_{-C} = 1.0\ s^{-1}$. The rate of attachments is now $k_+ = 1.8\ s^{-1}$ as both species make attempts to join the chain, whereas detachments continue at $k_- = 1.0\ s^{-1}$ because at any

Author contributions: C.J. wrote the paper.

The author declares no conflict of interest.

See companion article on page 9516.

*E-mail: cjarzyns@umd.edu.

instant only the monomer at the end can detach. Under these conditions the polymer grows, because $k_+ > k_-$. To make the connection with Eq. **2**, we note that while the thermodynamic force remains negative, $\varepsilon = -g_A/T = -g_C/T = \ln 0.9 \approx -0.1$, the disorder is $D = \ln 2 \approx 0.7$ for a random sequence of *A*s and *C*s in equal portions. Therefore, $\varepsilon + D \approx +0.6$, and the condition for polymer growth stipulated by Eq. **2** is satisfied. In this scenario, the polymer grows at a steady rate because the increase in free enthalpy that comes with each added monomer, g, is more than offset by the disordered information written into the sequence ω.

Eq. **2** reveals sequence randomness to be "a positive driving force favoring copolymer's growth" (6). The greater the disorder, the stronger the force. This echoes a similar result for information erasure: The greater the randomness in a sequence 011010$\cdots$, the more work is needed to set it to 000000$\cdots$. This raises a potentially interesting question: What if the polymer sequence is not necessarily random, but complex, like the digits of π? The work required to erase a complex pattern of 0s and 1s is determined by its algorithmic complexity (8), the length of the shortest set of instructions that generate the pattern. For complex but not necessarily random polymers, should sequence randomness be replaced by algorithmic complexity in Eq. **2**?

Andrieux and Gaspard (6) next consider copolymerization in the presence of a template, a blueprint encoding instructions to generate the polymer. For DNA replication, this template is itself a sequence of nucleotides, to be copied according to the base-pairing rule, *A*–*T* and *C*–*G*. As in the case without a template, the focus is on steady polymer growth, with average free enthalpy per monomer, g, and thermodynamic force, $\varepsilon = -g/T$. The analysis here is slightly more complicated than for the case without a template, and they obtain the following condition for steady polymer growth:

$$\varepsilon + D(\text{polymer}) - I(\text{polymer, template}) > 0. \quad \textbf{[3]}$$

Here, I(polymer, template) is the mutual information, per monomer, between the template sequence and the polymer sequence. This quantity measures the degree to which knowledge of one sequence specifies the other, and for DNA replication it achieves its maximum value, $I = D$, when the base pairs are copied with no errors. Eq. **3** thus tells us that the more faithfully the polymer is copied from the template, the greater the minimum thermodynamic force needed to keep the chain growing at a steady pace. In particular, copying with perfect fidelity ($I = D$) necessarily requires a positive thermodynamic driving force, $\varepsilon > 0$; whereas if some errors are tolerated, then we can in principle get away with a negative driving force, as long as $\varepsilon + D - I$ is positive.

These very general considerations are illustrated with model numerical simulations of DNA replication, in which the thermodynamic force ε is a control parameter. Several quantities, including polymer growth rate, fraction of copying errors, and mutual information, are investigated as functions of this force. The results illustrate, among other things, that the fraction of copying errors decreases as the thermodynamic force is turned up. Two distinct regimes are evident: (*i*) at small or slightly negative values of ε, a regime of low dissipation and high error rate, in which polymer growth is driven primarily by sequence randomness; and (*ii*) at large ε, a regime of high dissipation and few copying errors, driven by free enthalpy. These results suggest that "polymerases must operate far from equilibrium to achieve a high fidelity" (6).

Although Andrieux and Gaspard (6) are not the first to consider the thermodynamics of biopolymer replication [see, for instance, Hopfield's model of kinetic proofreading (9)], their fundamental results elegantly clarify and quantify the role of sequence randomness as a thermodynamic force in its own right. In doing so, these results strengthen an important link between two hallmark properties of biological systems (10): the information processing at the heart of self-replication, and the dissipation of energy inherent to metabolism.

1. Maxwell JC (1871) *Theory of Heat* (Longmans, Green, and Co, London); reprinted (2001) (Dover, New York).
2. Szilard L (1929) Über die Entropieverminderung in einem Thermodynamischen System bein Eingriffen Intelligenter Wesen. *Z Phys* 53:840–856.
3. Brillouin L (1951) Maxwell's demon cannot operate: Information and entropy. I. *J Appl Phys* 22:334–337.
4. Landauer R (1961) Irreversibility and heat generation in the computing process. *Adv Chem Phys* 5:183–191.
5. Bennett CH (1973) Logical reversibility of computation. *IBM J Res Dev* 17:525–532.
6. Andrieux D, Gaspard P (2008) Nonequilibrium generation of information in copolymerization processes. *Proc Natl Acad Sci USA* 105:9516–9521.
7. Cover TM, Thomas JA (1991) *Elements of Information Theory* (Wiley, New York).
8. Zurek WH (1989) Thermodynamic cost of computation, algorithmic complexity and the information metric. *Nature* 341:119–124.
9. 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.
10. Schrödinger E (1944) *What is Life?* (Cambridge Univ Press, Cambridge, UK).