NONEQUILIBRIUM THERMODYNAMICS AND THE ORIGIN OF LIFE

Gerald E. Marsh
Argonne National Laboratory (Retired)
gemarsh@uchicago.edu

ABSTRACT

Modern developments in nonequilibrium thermodynamics have significant implications for the origins of life. The reasons for this are closely related to a generalized version of the second law of thermodynamics recently found for entropy production during irreversible evolution of a given system such as self-replicating RNA. This paper is intended to serve as an introduction to these developments.

Introduction

It is generally believed that life began with the evolution of self-replicating polynucleotides. Ever since the paper by Gilbert[1] that discussed the possibility that catalytic RNA enzymes or ribozymes could be involved in the evolution of life, and incidentally coined the term the "RNA world", RNA has been the favorite molecule. Gilbert's *News and Views* paper addressed the previous week's *News and Views* paper by Westheimer[2]. The original discovery of the enzymatic activities of RNA was by Cech[3]. In 2009, Lincoln and Joyce[4] showed the self-sustained replication of two ribozymes that catalyze each other's synthesis. These cross-replicating ribozymes grew exponentially in the absence of proteins or other biological materials.

RNA molecules are composed of purine and pyrimidine nucleotides. How these nucleotides could have formed concurrently under the geophysical constraints of the early Earth was an unsolved chemical mystery before Becker, et al.[5] published their paper "Unified prebiotic plausible synthesis of pyrimidine and purine RNA ribonucleotides" in 2019. They found a reaction network under which both nucleotides could simultaneously form driven by wet-dry cycles. The chemical reactions involved are very complex, but Hud and Fialho[6] have summarized them as shown in Figure 1.
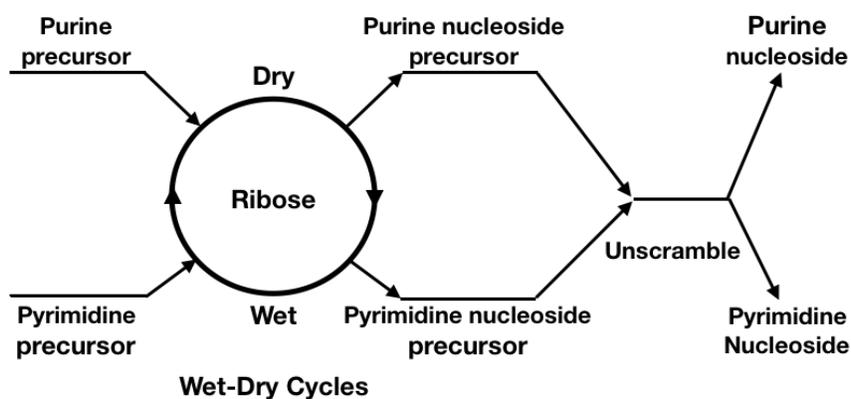

Figure 1. A simplified reaction network for the simultaneous formation of purine and pyrimidine nucleosides. When a mixture of the sugar ribose and non-natural nucleobases are subjected to wet-dry cycles, intermediate molecules are formed that can be converted to the natural nucleosides, which when phosphorylated give the natural nucleotides. For a discussion of what "scrambled" means see the article by Hud and Fialho from which this figure is adapted. They used the term to designate the precursor formed by Becker, et al., which they called a "scrambled" pyrimidine nucleobase.

The geochemical conditions in the current work of Becker, et al. is compatible with their previous synthesis of the purine nucleosides[7]. The wet-dry cycles allow the dried reactants to coalesce into a concentrated state where their joining through covalent bond formation with the release of water molecules becomes thermodynamically favorable. What "thermodynamically favorable" means is the subject of much of what follows.

From a simplistic point of view, life appears to violate the second law of thermodynamics since ordered life forms have a lower entropy than their precursors. Of course, this apparent violation vanishes when one realizes that life processes draw upon the free energy from the surroundings so that total entropy increases. In terms of the origin of life, it has been known for some time that matter can spontaneously organize itself if the Gibbs free energy for the process is negative or if an external source of "activation energy" is provided. If this absorbed energy is dissipated after overcoming the activation barrier it is no longer available to drive the reverse process. This irreversible phenomenon is called "driven self-assembly". Explaining how this works for a macroscopic system requires the introduction of a generalized form of the second law of thermodynamics.

A short Appendix discusses some of the terminology, definitions, and confusions associated with thermodynamic quantities.

Nonequilibrium States

A system may be maintained in a nonequilibrium state by a flow of energy. If the state is time independent, macroscopic observables will have constant nonequilibrium values. The example often given is a constant electrical current through a resistor with a steady rate of heat generation. Such dissipative structures can be formed and maintained by irreversible processes that continuously increase entropy.[8,9] In a linear regime, small deviations in the forces driving a flow (such as the heat flow on the resistor example) will lead to the flow being a linear function of the forces driving the flow. But a system that is not in thermodynamic equilibrium need not be in a stationary, time independent state since systems that are far from equilibrium can become dependent on nonlinear phenomenological laws.

Systems of identical non-interacting components in thermodynamic equilibrium can be described by the Boltzmann distribution (also known as the Gibbs distribution), which gives the equilibrium probability distribution of different energy states of a system as a function of the state's energy and the temperature of the system. It has the general form $P_i \propto \exp[-E_i/k_B T]$, where k_B is the Boltzmann constant, P_i is the probability that the system is in the state i and E_i is the energy of the state. Thus, if the states j and k have energies E_i and E_j, the relative probability is

$$\frac{P_i}{P_j} = \exp\left[\frac{E_j - E_i}{k_B T}\right].$$

(1)

Moreover, systems in thermodynamic equilibrium satisfy the principle of detailed balance, which is equivalent to microscopic reversibility.

If the system is driven far from equilibrium into the nonlinear regime, the Boltzmann distribution is no longer valid since the thermodynamic flows are no longer linear functions of the thermodynamic forces. Far from equilibrium states can evolve into one of many new, highly organized states also known as dissipative structures.

Microstates and Reversibility

The statistical behavior of nonequilibrium systems require the introduction of comparisons between the dynamical trajectories of the components of the system rather than the local properties of individual microstates at one moment of time.[10] If the dynamics of a system are stochastic and Markovian (meaning a sequence of events where the probability of each event depends only on the state of the previous event), one can require that the dynamics follow the microscopically reversible condition[11]

$$\frac{P[x(+t)|\lambda(+t)]}{P[\bar{x}(-t)|\bar{\lambda}(-t)]} = exp\{-\beta\ Q[x(+t),\lambda(+t)]\},$$

(2)

4

where $\beta = 1/k_B T$, the state of the system is given by the function x, representing all dynamical uncontrolled degrees of freedom, while λ is a controlled time-dependent parameter.

In Eq. (2), $P[x(+t)|\lambda(+t)]$ is the probability of following the path $x(+t)$ through phase space and the denominator is the corresponding time-reversed path. This notation for the time-reversed path is a consequence of changing the time origin so that $t \in \{-\tau, \tau\}$, where τ could be infinite. The overbar indicates that quantities odd under time reversal also change sign. Q is the amount of energy in the form of heat transferred to the system from the heat bath. Q is a function of the phase space path and odd under time reversal; i.e., $Q[x(+t), \lambda(+t)] = -Q[\bar{x}(-t), \bar{\lambda}(-t)]$.

Eq. (2) is microscopically reversible; it relates the probability of a given path to its reverse path. Note that this is not the same as the principle of detailed balance, which refers to the probabilities of changing states independent of path. This distinction is important because Eq. (2) holds when the system is driven by an external time varying force field. English[12] has used this equation to derive a generalization of the second law of thermodynamics that is important for many far from equilibrium thermodynamic systems. It applies to the macroscopic transition between complex course-grained states.

Dissipative Adaptation

To be consistent with English's notation, rewrite Eq. (2) as

$$\frac{\pi[\gamma]}{\pi^*(\gamma^*)} = exp\left[\frac{\Delta Q(\gamma)}{kT}\right],$$

(3)

where γ is a microtrajectory and the "$*$" indicates time-reversed. Here, it is assumed that $\Delta Q(\gamma)$ is the sum of the internal energy change of the system when traversing the path γ and the work applied to the system by an external field; i.e., $\Delta Q(\gamma) = \Delta E + W$.

Let i, j, and k represent different possible configurations of a system composed of distinct components (such as particles). Equation (3) then becomes

$$\frac{\pi[i \to j]}{\pi^*(j^* \to i^*)} = exp\left[\frac{\Delta Q(i \to j)}{kT}\right].$$

(4)

Using Eq. (4) also for the trajectory $i \to k$ and taking the ratio of the equations for the transition $i \to j$ and $i \to k$, that is,

$$\frac{\frac{\pi[i \to j]}{\pi^*(j^* \to i^*)}}{\frac{\pi[i \to k]}{\pi^*(k^* \to i^*)}} = \frac{exp\left[\frac{\Delta Q(i \to j)}{kT}\right]}{exp\left[\frac{\Delta Q(i \to k)}{kT}\right]},$$

(5)

with the substitution $\Delta Q(\gamma) = \Delta E + W$, after some algebra Eq. (5) will yield

$$\frac{\pi[i \to j]}{\pi[i \to k]} = exp\left[\frac{-\Delta E(j \to k)}{kT}\right]\frac{\pi^*(j^* \to i^*)}{\pi^*(k^* \to i^*)}\frac{exp\left[-\frac{W(i \to k)}{kT}\right]}{exp\left[-\frac{W(i \to j)}{kT}\right]}.$$

(6)

The first term on the right-hand-side of Eq. (6) comes from the term $exp\left[\frac{\Delta E(i \to j) - \Delta E(i \to k)}{kT}\right]$ encountered when doing the algebra. The energy level relationships of Eq. (6) are shown in the figure below
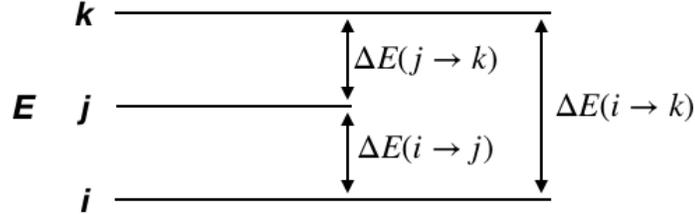
From this figure one can see that $\Delta E(i \to j) - \Delta E(i \to k) = \Delta E(j \to k)$.

To explain the concept of dissipative adaptation, English assumes that the states have the same energy so that the first term on the right-hand side of Eq. (6) is unity, and averages over all microtrajectories with fixed endpoints. With these assumptions, Eq. (6) becomes

$$\frac{\pi[i \to j]}{\pi[i \to k]} = \frac{\pi^*(j^* \to i^*)}{\pi^*(k^* \to i^*)}\frac{exp\langle-\frac{W(i \to k)}{kT}\rangle}{exp\langle-\frac{W(i \to j)}{kT}\rangle}.$$

(7)

The brackets $\langle \dots \rangle$ indicate the average over microtrajectories. Note that even though the states now have the same energy, the left-hand side of Eq. (7) could differ from unity since not all states of the system are equally accessible in a finite time.

For different oscillatory external forces, known as the "drive", different configurations of the system will absorb work from the external forces at different rates. A given system configuration could then surmount "activation barriers" to transition to states that would not be accessible through thermal fluctuations alone. This is shown in Fig. 2 for the general case where the energies of the states are not identical.
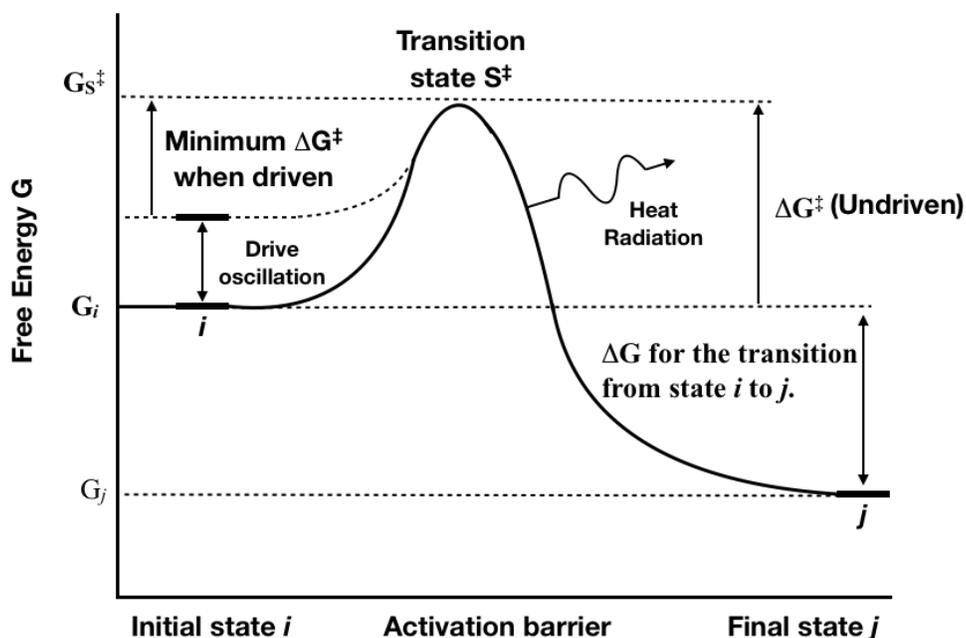

Figure 2. Transition through a system configuration activation barrier. $\Delta G^{\ddagger} = G_S^{\ddagger} - G_i$, where G_i oscillates when driven. For the case where the energies of the states i and j are equal, $G_i = G_j$. The energy absorbed from the drive is radiated as heat in the transition from $S^{\ddagger}$ to state j. This figure can be compared to Fig. A1 in the Appendix for the activation energy of a chemical reaction.

The point of all this is that some of the randomly changing system configurations will be better able to absorb work from the drive than others and this leads to a mostly one directional change in configurations—due to the loss of the radiated heat energy—to those better able to absorb and dissipate the energy absorbed from the drive. As put by English, the structure will appear to self-organize into a state that is well adapted to the environmental conditions set by the drive. This, he calls the phenomenon of "dissipative adaptation".

In this vein, Kondepudi and Prigogine in their 2015 book *Modern Thermodynamics* (Ref. 8) introduce the concept of "structural instability" that can occur when a new chemical species is introduced into a nonequilibrium chemical system that could destabilize the system so that it evolves into a new state. They liken this to Darwinian evolution at the molecular level. Thus, in their words "we see instability, fluctuation and evolution to organized states as a general nonequilibrium process whose most spectacular manifestation is the evolution of life."

Macroscopic Irreversibility

The relationship between microscopic irreversibility and entropy production has thermodynamic effects on far from equilibrium macroscopic processes and in particular for biological self-replication (to be discussed in the next section). English derived a generalized form of the second law of thermodynamics that can be written as

$$\beta \langle \Delta Q \rangle_{I \to II} + ln \left[\frac{\pi(II \to I)}{\pi(I \to II)} \right] + \Delta S_{int} \geq 0.$$

(8)

The derivation of this equation begins with Eq. (3) with $\gamma \to x(t)$, where $0 \leq t \leq \tau$. Taking the natural logarithm of the resulting equation gives

$$\beta \Delta Q = ln \left[\frac{\pi[x(t)]}{\pi[x(\tau - t)]} \right],$$

(9)

an equation concerned with microscopic irreversibility.

Suppose there is a coarse-grained observable, *I*, which can be associated with a probability distribution $p(i|I)$, the probability that it is in a microstate *i*. If there is a second course-grained observation of the system after a time interval τ designated by *II*, $p(j|II)$ is defined as the probability that the macrostate *II* (that evolved from macrostate *I* after a period of time τ) is in the microstate *j*. Note that the macrostates I and II are ensembles of paths.

Crucially, these probability functions allow a macroscopic definition of irreversibility:

$$\pi(I \to II) = \int_{II} dj \int_{I} di \, p(i|I) \, \pi(i \to j)$$

$$\pi(II \to I) = \int_{I} di \int_{II} dj \, p(j|II) \, \pi(j \to i).$$

(10)

Similar to what was done in Eq. (5), taking the ratio of $\frac{\pi(II \to I)}{\pi(I \to II)}$ gives, after some algebra,

$$\frac{\pi(II \to I)}{\pi(I \to II)} = \langle \, \langle e^{-\beta \, \Delta Q_{ij}} \rangle_{i \to j} / \exp \left[ln \left[\frac{p(i|I)}{p(j|II)} \right] \right] \, \rangle_{I \to II}.$$

(11)

The first averaging bracket on the right-hand side is the average of all paths from $i \in I$ to $j \in II$, each path being weighted by its likelihood (the second averaging bracket).

The next step is to introduce the Shannon entropy $S = -\sum_i p_i \, ln \, p_i$ so that an expression for the internal entropy change for the transition between the ensembles $I \to II$ can be written. English uses units such that the Boltzmann constant is unity. This results in $\Delta S_{int} = S_{II} - S_I$ and after some additional algebra he obtains the generalization of the second law of thermodynamics given by Eq. (8); that is,

$$\beta \langle \Delta Q \rangle_{I \to II} + ln \left[\frac{\pi(II \to I)}{\pi(I \to II)} \right] + \Delta S_{int} \geq 0.$$

(12)

The first term in Eq. (12) is the entropy change of the heat bath and the entropy generated by second term vanishes if $\pi(II \to I) = \pi(I \to II)$, which would result in the usual second law of thermodynamics where $\beta \langle \Delta Q \rangle_{I \to II} + \Delta S_{int} \geq 0$ because the average entropy change of the universe must be greater than or equal to zero.

One might question whether the use of the Shannon entropy is legitimate. The Boltzmann distribution implies the usual thermodynamic definition of entropy. The Gibbs-Shannon entropy

given by $S = -k_B \sum_i p_i \ln p_i$ is equivalent to the thermodynamic definition of entropy only for what is known as the generalized Boltzmann distribution,[13] which is valid for all Markovian systems even those not in thermodynamic equilibrium. The generalized Boltzmann distribution itself was defined by Lin[14] using an analogy based on electronics to give an explanation of the concept. Given a thermodynamic system with $m + n$ generalized forces and coordinates, Xiang, et al. write the probability density function $Pr(\vec{\omega})$ of the microstate $\vec{\omega}$ for the generalized Boltzmann distribution as

$$Pr(\vec{\omega}) \propto exp\left[\sum_{\eta=1}^{n} \frac{X_\eta x_\eta^{(\vec{\omega})}}{k_B T} - \frac{E^{(\vec{\omega})}}{k_B T}\right],$$

(13)

where the X_η are generalized forces and E and x_η are random variables, and Xiang, et al. use the vector notation to designate a microstate, as in $\vec{\omega}$. If the generalized forces vanish, Eq. (13) reduces to the usual Boltzmann distribution.

Since the systems considered above are Markovian, the use of the Shannon entropy is indeed legitimate.

Replicating Systems and the Generalized Second Law of Thermodynamics

English has used self-replicating systems to illustrate the use of the generalized second law. Consider $n(t = 0) \gg 1$ for n self-replicating molecules at an inverse temperature β. They would have an exponential growth given by

$$n(t) = n(0)e^{(g-\delta)t},$$

(14)

where g determines the growth rate and δ the decay rate.

The probability that in a time dt one particular replicator associated with $\pi(I \rightarrow II)$ would replicate would be given by $g dt$ and its decay probability $\pi(II \rightarrow I)$ would be δdt. Imposing the generalized second law of thermodynamics yields,

$$\Delta S_{tot} = \beta \Delta q + \Delta S_{int} \geq ln\left[\frac{g}{\delta}\right].$$

(15)

Note that if $g > \delta$, so that there is net growth, the total entropy associated with self-replication will have a positive lower bound.

Two conclusions are readily apparent: (1) The growth rate of a self-replicator depends on its internal entropy (ΔS_{int}), its durability ($1/\delta$), and the heat (Δq) dissipated into the surrounding heat bath in the process of replication; and (2), heat must be generated from energy stored in the reactants or work done on the system by a time varying external driving field.

Implications for the Origin of Life

Using the RNA molecule again as an example, how purine and pyrimidine nucleosides could have formed together under early Earth geophysical constraints was, as mentioned in the Introduction, until recently an unsolved chemical problem. From a global perspective, the resolution of this problem involved far from equilibrium thermodynamics—in this case where the system was driven by wet-dry cycles; it is this external forcing that made the formation of these nucleosides thermodynamically favorable.

A very interesting mechanism has been proposed by K. Michaelian[15] for the replication of RNA and DNA molecules without the need for enzymes. He proposed that UV light dissipation and diurnal temperature cycling of the Archean sea-surface can lead to such replication. It is based on what Michaelian calls a non-equilibrium thermodynamic imperative for producing RNA and DNA due to the great entropy producing potential of these molecules under the initial conditions of the primitive Earth.

While the idea that far from equilibrium thermodynamics is fundamental to the origin of life remains somewhat controversial, the case is quite strong for this point of view. There is also a simulated toy chemical model whose behavior is consistent with the idea of far from equilibrium self-organization.[16]

General chemistry tells us that the free energy change in a reaction is governed by the equilibrium constant. Furthermore, while the reaction rate can be changed by enzymes which do not alter the equilibrium, the change in free energy is independent of the path or the molecular mechanism of the transformation. It is clear that this is not the case for nonequilibrium thermodynamics "whose most spectacular manifestation is the evolution of life".

Our Earth formed some 4.5 billion years ago and there is strong evidence that life appeared 700 million to 1 billion years later; that is, about 3.8 billion years ago. Far from equilibrium nonlinear thermodynamics in the presence of external drives could help explain this very rapid origin of life. Because the length of the day was only around seven or so hours long 3.8 billion years ago, the tides would be far higher than today and there would be strong diurnal forces that could play the role of an external drive including UV radiation, which would be far more intense than today, as pointed out by Michaelian, since there was little if any oxygen in the atmosphere and therefore no ozone layer to block the UV.

Life appears as a process ultimately founded on the ability of matter, governed by the principles of quantum mechanics, to form the molecules need for life to exist. These molecules combine, overcoming the limits normally imposed by both unfavorable free energy constraints and activation energies, because of the properties of driven nonlinear thermodynamic systems. Most stars have planets, and those with earthlike planets are all very likely to have life due to the reduction in the activation energy for the formation of complex biomolecules arising from driven, far from equilibrium nonlinear thermodynamics.

Acknowledgement

APPENDIX

There is much confusion in the literature about the term "energy" in physics, biology, and chemistry. The definitions here are from the International Union of Pure and Applied Chemistry (IUPAC). They define:

U = the internal energy, $U = Q + W$, where Q designates heat and W work. In parts of this paper, to be consistent with the referenced literature, the internal energy is designated as E and its change ΔE.

A = Helmholtz energy function: $A = U - TS$, where S is the entropy and T the temperature. Note that in the earlier literature A was called the "free energy".

H = enthalpy: $H = U + PV$, where P is the pressure and V the volume. ΔH is the heat brought to a system at constant pressure.

G = Gibbs energy function: $G = H - TS$. Formerly called the "free energy" or "free enthalpy". It is the reversible useful work performed at constant temperature and pressure.

An additional confusion is that in physics the "free energy" is the generally the Helmholtz energy and in biology and chemistry it is the Gibbs free energy. In particular, in biochemistry, the Gibbs free energy is defined as

$$\Delta G = \Delta G^0 + RT \, ln \frac{[C][D]}{[A][B]},$$

where, $A + B \leftrightharpoons C + D$, and $[A, B, C$ or $D]$ are the molar concentrations (the "activities") of the reactants. ΔG^0 is the Gibbs free energy for this reaction under "standard conditions", meaning when A, B, C, D are present at a concentration of 1.0 M.

In biochemistry, it is generally assumed that the initial and final states of a reaction are equilibrium states so that $\Delta G = G_{final} - G_{initial}$. ΔG is independent of the path or molecular mechanism of the transformation. This is not the case for the nonequilibrium states discussed in the body of this paper. A negative ΔG implies that the reaction can occur spontaneously. The rate of a reaction depends on the "free energy of activation" $\Delta G^{\ddagger}$ as shown in the figure below.
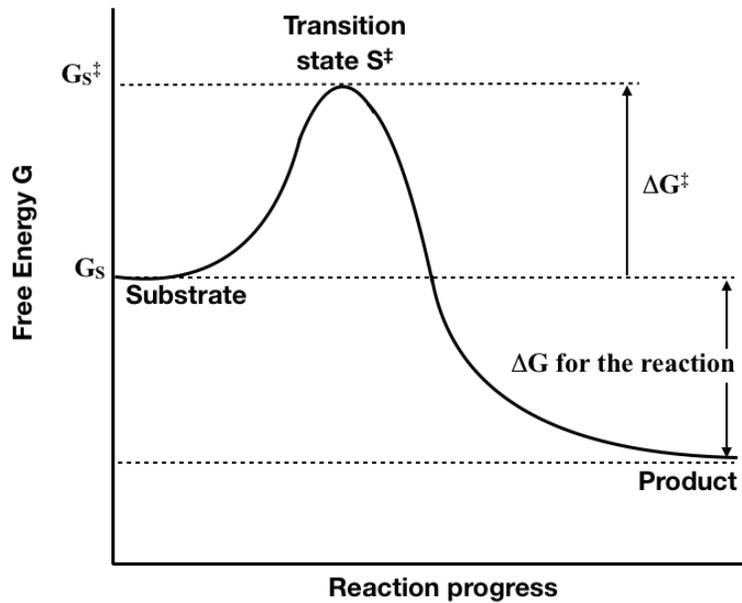

Figure A1. The transition state S‡ has a higher energy than either the substrate S or product. The Gibbs free energy of activation, or "activation energy", is $G^{\ddagger} = G_{S^{\ddagger}} - G_S$, where $G_{S^{\ddagger}}$ is the Gibbs free energy of the transition state S‡, and G_S is the free energy of the substrate.

REFERENCES

[1] W. Gilbert, "The RNA world", *Nature* **319**, 618 (20 February 1986).

[2] F.H. Westheimer, "Biochemistry: Polyribonucleic acids as enzymes", *Nature* **319**, pp. 534-536 (13 February 1986).

[3] T.R. Cech, "Self-Splicing RNA: Implications for Evolution", *Int. Rev. Cytol.* **93**, pp. 3-22 (1985).

[4] T.A. Lincoln and G.F Joyce, "Self-Sustained Replication of an RNA Enzyme", *Science* **323**, pp. 1229-1232 (27 February 2009).

[5] S. Becker, et al., Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides", *Science* **366**, pp. 76-82 (4 October 2019).

[6] N.V. Hud and D.M. Fialho, "RNA nucleosides built in one prebiotic pot", *Science* **366**, pp. 32-33 (4 October 2019).

[7] S. Becker, et al., "A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway", *Science* **352**, pp. 833-836 (13 May 2016).

[8] D. Kondepudi and I. Prigogine, *Modern Thermodynamics: From Heat Engines to Dissipative Structures* (John Wiley & Sons, Ltd., United Kingdom 2015).

[9] B. DeBari, et al., "Oscillatory dynamics of an electrically driven dissipative structure", PLOS ONE | https://doi.org/journal.pone.0217305 (May 29, 2019).

[10] J.L. English, "Dissipative adaptation in driven self-assembly", *Nature Nanotechnology* **10**, pp. 919-923 (November 2015).

[11] G.E. Crooks, "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences", *Phys. Rev.* E **60**, pp. 2721-2726 (1999).

[12] J.L. English, "Statistical physics of self-replication" *J. Chem. Phys.* **139**, pp. 121923 to 1-121923-8 (2013).

[13] Xiang Gao, et al., "The Generalized Boltzmann Distribution is the Only Distribution in Which the Gibbs-Shannon Entropy Equals the Thermodynamic Entropy", arXiv:1903.02121 [cond-mat.stat-mech] (12 Apr 2019).

[14] M.M. Lin, "Generalized Boltzmann distribution for systems out of equilibrium", arXiv:1610.02612 [cond-mat.stat-mech] (13 Oct 2018).

[15] K. Michaelian, "Thermodynamic dissipation theory for the origin of life", *Earth Syst. Dynam.*, **2**, pp. 37–51 (2011). Interchanges between the author and referees are available at https://www.earth-syst-dynam.net/2/37/2011/esd-2-37-2011-discussion.html#discussion.

[16] T. Kachman, J.A. Owen, and J.L. England, "Self-Organized Resonance during Search of a Diverse Chemical Space", *Phys. Rev. Lett.*, **119**, pp. 038001-1 to 038001-5 (2017).