

RESEARCH ARTICLE

Investigating Information Dynamics in Living Systems through the Structure and Function of Enzymes

Robert Gatenby^{1*}, B. Roy Frieden²

1 Departments of Integrated Mathematical Oncology and Radiology, Moffitt Cancer Center, Tampa, FL 33612, United States of America, **2** College of Optical Sciences, University of Arizona, Tucson, AZ 85721, United States of America

* Robert.Gatenby@moffitt.org



CrossMark
click for updates

OPEN ACCESS

Citation: Gatenby R, Frieden BR (2016) Investigating Information Dynamics in Living Systems through the Structure and Function of Enzymes. PLoS ONE 11 (5): e0154867. doi:10.1371/journal.pone.0154867

Editor: Eugene A. Permyakov, Russian Academy of Sciences, Institute for Biological Instrumentation, RUSSIAN FEDERATION

Received: February 24, 2016

Accepted: April 20, 2016

Published: May 5, 2016

Copyright: © 2016 Gatenby, Frieden. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper.

Funding: The authors acknowledge support from the National Cancer Institute under grant NIH/NCI 1 U54CA193489-01A1. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Enzymes are proteins that accelerate intracellular chemical reactions often by factors of 10^5 – 10^{12}s^{-1} . We propose the structure and function of enzymes represent the thermodynamic expression of heritable information encoded in DNA with post-translational modifications that reflect intra- and extra-cellular environmental inputs. The 3 dimensional shape of the protein, determined by the genetically-specified amino acid sequence and post translational modifications, permits geometric interactions with substrate molecules traditionally described by the key-lock best fit model. Here we apply Kullback-Leibler (K-L) divergence as metric of this geometric “fit” and the information content of the interactions. When the K-L ‘distance’ between interspersed substrate p_n and enzyme r_n positions is *minimized*, the information state, reaction probability, and reaction rate are *maximized*. The latter obeys the Arrhenius equation, which we show can be derived from the geometrical principle of minimum K-L distance. The derivation is first limited to optimum substrate positions for fixed sets of enzyme positions. However, maximally improving the key/lock fit, called ‘induced fit,’ requires *both* sets of positions to be varied optimally. We demonstrate this permits and is maximally efficient if the key and lock particles p_n, r_n are quantum entangled because the level of entanglement obeys the same *minimized* value of the Kullback-Leibler distance that occurs when all $p_n \approx r_n$. This implies interchanges $p_n \rightleftharpoons br_n$ randomly taking place during a reaction successively improves key/lock fits, reducing the activation energy E_a and increasing the reaction rate k . Our results demonstrate the summation of heritable and environmental information that determines the enzyme spatial configuration, by decreasing the K-L divergence, is converted to thermodynamic work by reducing E_a and increasing k of intracellular reactions. Macroscopically, enzyme information increases the order in living systems, similar to the Maxwell demon *gedanken*, by selectively accelerating specific reaction thus generating both spatial and temporal concentration gradients.

Introduction

Living organisms, uniquely in nature, encode, propagate, and use information [1] to produce stable, highly-ordered structures that are also complex, dynamical, semi-open systems far from thermodynamic equilibrium. But, what is biological information and how is information used to maintain the ordered structure and function of a living system [2–5]? While it is apparent that information storage and process are fundamental characteristics of living systems, the principles governing information dynamics in biology remain unclear.

Enzymes are central to the function of living systems and facilitate the work necessary to maintain order [6]. Once synthesized as a string of amino acids specified by the nucleotide triplets in the gene, a protein is typically subjected to post-translational modification such as phosphorylation. Importantly, post translational modifications reflect temporally variations in the status of the cell (e.g. ATP concentrations [7]). Thus the 3 dimensional shape of the enzyme represents a summation of both heritable and current information within the cell. This composite information produces a 3 dimensional structure that is the low free-energy state for the amino acid sequence plus post-translational modifications. It will be seen that this minimum state represents, as well, one of minimum Kullback-Leibler divergence, i.e. maximal order, between substrate and enzyme codons. These effects result from a doubly-optimized lock and key interaction between substrate and enzyme codons.

By this effect, the enzymes are catalysts that do not alter the fundamental thermodynamics of the reaction, in the sense that the initial thermodynamic state of substrate and the final thermodynamic state of the products are not changed [8]. Because it acts as a catalyst, the enzyme is not consumed in the reaction so that its information content is applied repeatedly provided substrate is available and no additional post-translation modifications occur. Typically enzymes accelerate reactions, often by many orders of magnitude (Fig 1). Without them, many reactions—e.g., reactions to extract energy from substrate or synthesize cell components—would be too slow to permit orderly function of living systems. We propose that this characteristic of enzymes permits investigation of the relation of information to thermodynamics and order through the concept of “activation energy.” Finally, we note recent studies [8,9] have emphasized the dynamic nature of enzyme structure and the critical role of structural motion of the protein during catalysis. By integrating these dynamics into our model, we note that quantum effects may be observed.

Modeling Methods and Results

Key-Lock dynamics

Enzymes are typically highly specific, decreasing the activation energy (E_a) (Figs 1 and 2) and increasing the reaction rate (k) only for a small number of substrate molecules [9]. This link between E_a and k is typically described by the empirically derived Arrhenius equation (see below). The specific activity of the enzyme is often described as a “lock and key” [10] process in which some region of the folded protein provides a complementary geometric shape to that of the substrate [11,12] thus reducing the entropy of the interactions. We note that enthalpic interactions such as Coulomb interactions are also maximized as the distance between substrate and enzyme is decreased where, as noted in [12], “interactive enthalpy is estimated from the sum of electrostatic and van-der-Waals interactions.” This permits binding that facilitates the reaction often through complex intermediate transitory steps.

Here we will focus on the spatial interactions between enzyme and substrate. We will view the catalyzed reaction as a single step of substrate \rightarrow products (Figs 1 and 2) omitting for simplicity the transient intermediate steps. We initially assume an enzyme density law $r_n = r(x_n)$,

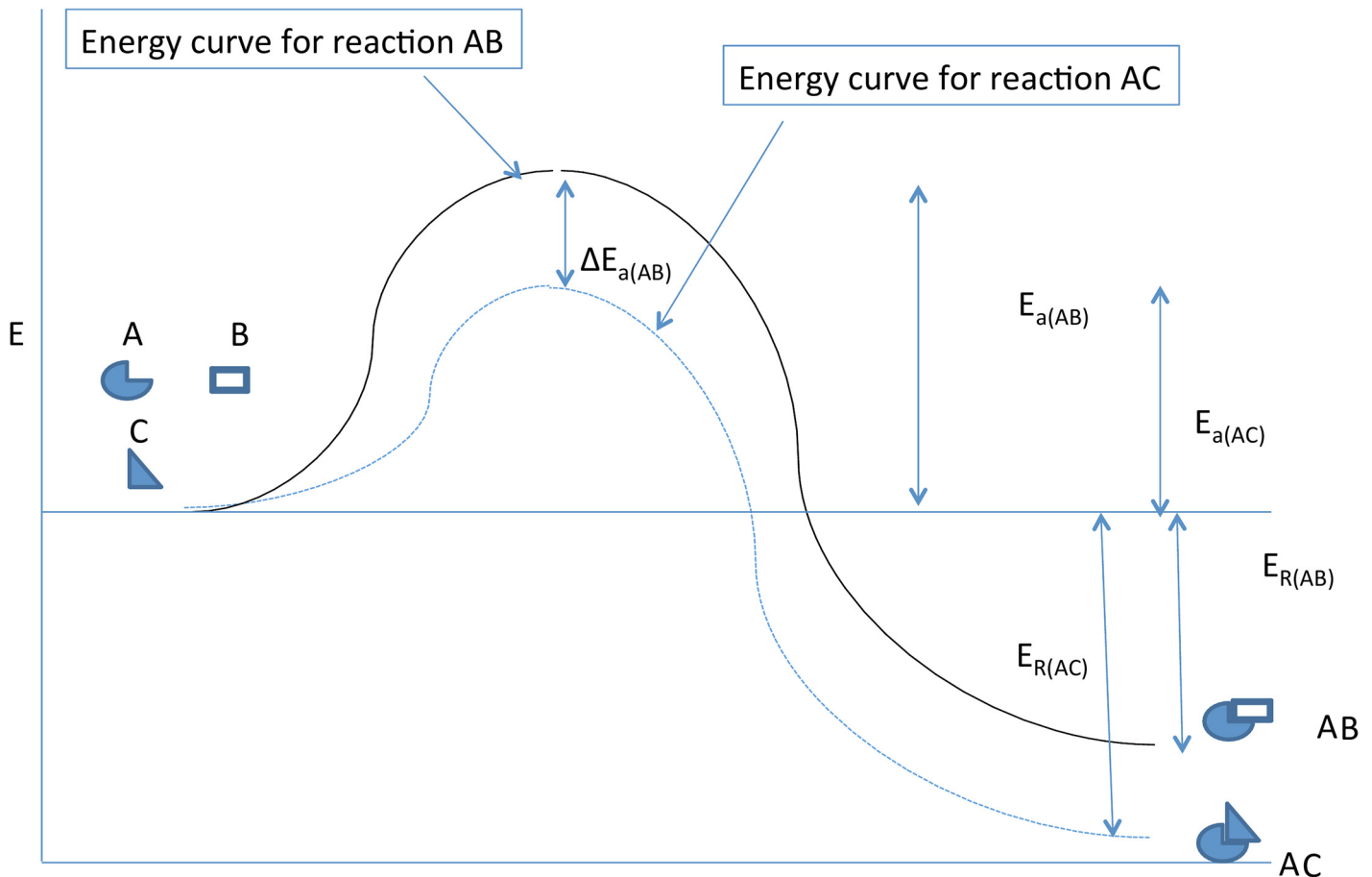


Fig 1. A simplified model of a reaction with and without an enzyme. Substrate B yields products C and D with a release of free energy E_R . Although the overall reaction is thermodynamically favorable, there is an energy barrier (the activation energy [E_a]) that decreases the rate of the reaction (k). The enzyme, through a key-lock geometric binding with the substrate, has a net effect of reducing the E_a and accelerating the reaction. As described in the text, the information content of the enzyme is expressed geometrically by the formation of a shape within the protein that is precisely complementary to the shape of the substrate. The information is, thus, converted to energy by reducing E_a (ΔE_a).

doi:10.1371/journal.pone.0154867.g001

$n = 1, \dots, N$ with the proteins fixed at molecular positions $x_n = n\Delta x$, $n = 1, \dots, N$. For simplicity, a one dimensional case is temporarily assumed, and with constant position spacings Δx . These constraints will be relaxed in subsequent sections.

Let the substrate pathway positions obey an unknown density law $p(x_n)$, $n = 1, \dots, N$ on the pairs of substrate particles that ordinarily constitute reactant molecules. Let these reactant molecules interact, or ‘bind,’ with the enzyme molecules. This defines an enzyme-substrate complex.

It is shown (see [Appendix](#)) that this complex lowers the activation energy of the reaction. One of the most important ways that an enzyme catalyzes any given reaction is through entropy reduction: by bringing order to a disordered system. Thus, since entropy is a component of Gibbs free energy, this free energy is lowered as well. This in turn is a component of the activation energy E_a which, as mentioned above, is likewise lowered. These factors work to increase the reaction rate. They also accelerate the reaction by providing a spatially specific charge distribution that form bonds with substrate to accelerate the reaction process. Enzymes also promote chemical reactions by bringing substrates together in an optimal orientation, lining up the atoms and bonds of one molecule with the atoms and bonds of the other molecule.

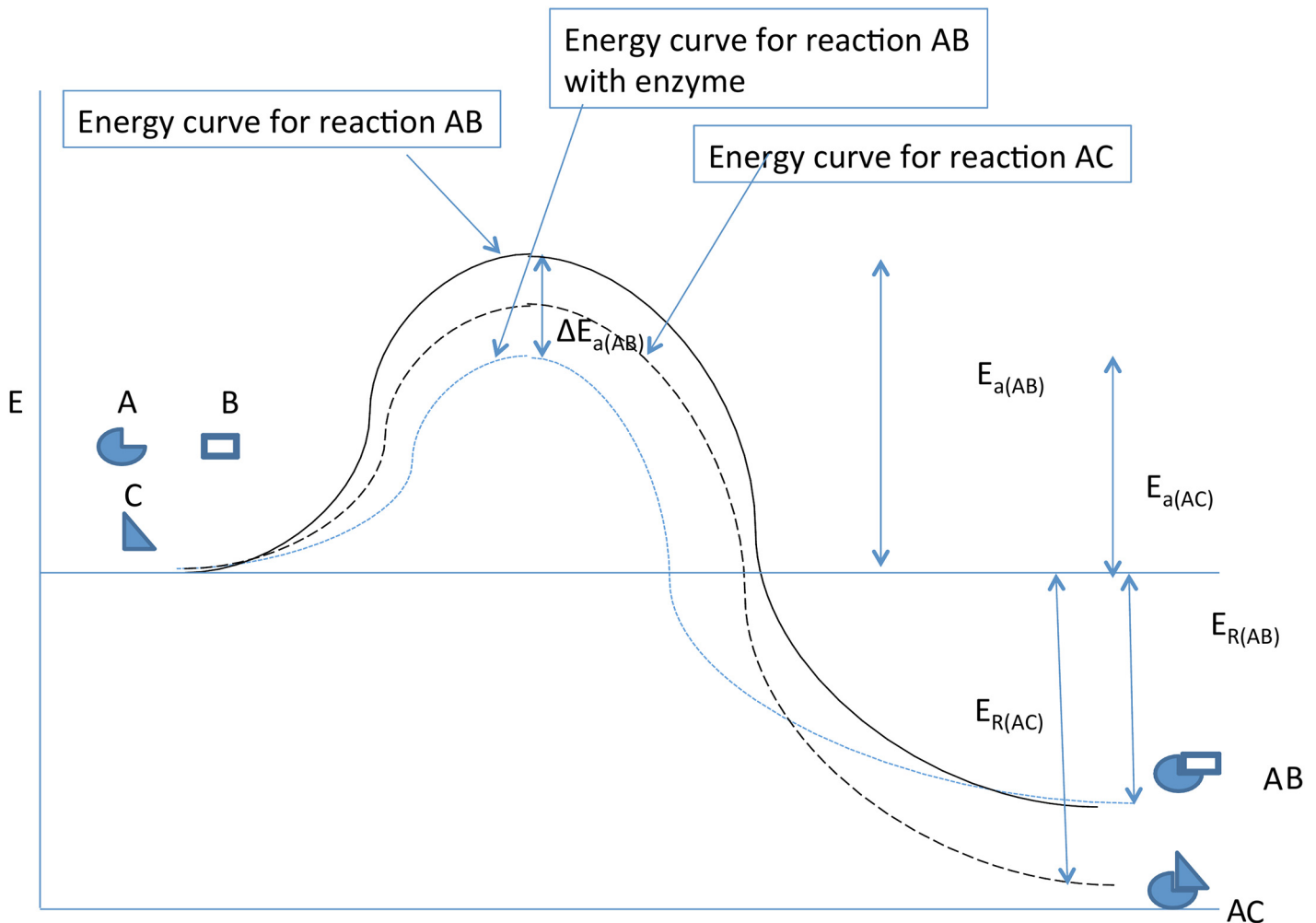


Fig 2. Information in living systems manifest through “temporal gradients”. Here the system contains initially two substrates and one enzyme. In the absence of the enzyme, reaction $C \rightarrow G + H$ will proceed more rapidly because it has both lower final free energy and lower activation energy. However, the enzyme lowers the E_a for reaction $B \rightarrow E + F$. The information in the enzyme produces an observable gradient over time as the concentrations of E and F are increased and B is decreased when compared to an uncatalyzed system. In contrast, because of its specificity, the enzyme has no effect on the temporal evolution of the substrate and product concentrations of reaction $C \rightarrow G + H$.

doi:10.1371/journal.pone.0154867.g002

This constitutes a lowering of local entropy, in particular the Kullback-Leibler or ‘cross’ entropy (as will be seen).

The initial interaction between enzyme and substrate is relatively weak, but these weak interactions rapidly induce conformational changes in the enzyme that strengthen binding [13]. These conformational changes are augmented by a ‘key and lock’ effect whereby the substrate ‘key’ molecule fits *optimally close* to the complementary 3 dimensional structure within the enzyme ‘lock’ particle. This ‘key/lock’ effect tends to maximize the reaction rate.

Initially assuming a well-mixed distribution of enzymes and substrate of equal concentration, we view the “lock” as constantly-spaced enzyme molecules of density profile $r_n = r(x_n + \Delta x/2)$, $x_n = n\Delta x$, Δx small. These molecules are located at positions $(n + 1/2)\Delta x$ with density values r_n . And by comparison, the substrate (or “key”) molecules are particle pairs having a local density profile $p_n = p(x_n)$ at positions $x_n = n\Delta x$. These are thereby located *halfway between* corresponding lock molecules r_n . Each enzyme-substrate ‘complex’ locally lowers the activation energy of the

reaction so that overall activation energy is maximally lowered when all key particles are ‘closest’ geometrically to the corresponding lock particles.

This is exemplified in Figs 1 and 2.

Then, given a fixed enzyme path r_n , the problem of minimizing activation energy becomes one of geometry. What substrate reactant path p_n obeys minimal distance from the fixed enzyme path $r(x_n)$?

Kullback-Leibler measure. We now need to choose a measure of the distance between the two density paths. From the preceding, this distance is to be a minimum. One useful measure is their Kullback-Leibler [14,15] ‘divergence,’ defined as

$$H_{KL}(p||r) = \sum_{n=1}^N p_n \ln \left[\frac{p_n}{r_n} \right]. \tag{1a}$$

Although H_{KL} is not formally a ‘distance’ (since it is not symmetric in p and r) it has many properties of one and, for our purposes, is convenient to be regarded as such. It also obviously has the form of an ‘entropy,’ and so can be termed ‘KL entropy’.

The KL distance between all enzymes of density $r_n = r(x_n + \Delta x/2)$ and their corresponding substrate molecules of density $p_n = p(x_n)$ is to be minimized, obeying

$$H_{KL}(p||r) = \min. \tag{1b}$$

We are here analyzing a one-dimensional problem, i.e. where each x_n and Δx is a scalar value. But this ignores the vital question of *relative orientation* of key and lock molecules. That is taken up at the end, and is an easy generalization of the one-dimensional approach.

This geometrical interleaving of the two types of molecule does represent a one-dimensional form of a key-lock geometry. However, specifically what *density function* $p(x_n)$ should govern the reactant pathway?

Derivation of optimum reactant pathway p_n . Regarding *all* enzyme and reactant molecules, this is assumed to obey principle (Eq 1a and Eq 1b). The reactant is also the substrate, so we are seeking the substrate *density function* p_n that has minimum KL distance from the *given* enzyme pathway r_n , $n = 1, \dots, N$. This is assumed to occur in the presence of the interlacing $(x_n, x_n + \Delta x/2)$ of coordinate positions defined above, and also the known physical constraints of the problem. The main one is that of known mean energy.

We seek the pathway position law p_n that obeys $H_{KL}(p||r) = \min.$, in the presence of the arbitrary, but fixed, enzyme pathway r_n . (Note: This temporarily ignores the more recently observed effect of “induced fit [16,17],” whereby the enzyme pathway changes as well to further improve the fit. This is addressed below. The two laws p_n, r_n of course obey normalization

$$\sum_n p_n = 1, \quad \sum_n r_n = 1, \quad r_n = \text{const.}, \quad n = 1, \dots, N. \tag{2}$$

(All sums are over the entire pathways.). Assume, as well, a fixed, *mean* molecular bond energy

$$\sum_n P(E_n)E_n = \sum_n p_n E_n = kT, \quad \text{with } P(E_n) = p_n \tag{3}$$

by definition, κ Boltzmann’s constant and T a fixed energy. Energies E_n could, e.g., be due to hydrogen bonds. Also, Eq (3) assumes ergodicity to hold. That is, the true statistical average energy—the left-hand sum—equals the average energy along any one path—the second sum. We will use this ergodic property below.

Net Optimization Problem. We therefore seek the reaction (or substrate) rate p_n satisfying KL requirement Eq (1b) subject to four constraints Eqs (2) and (3) obeyed by p_n and r_n . By

the method of undetermined multipliers, these satisfy the variational principle

$$\sum_n p_n \ln \left[\frac{p_n}{r_n} \right] + \Lambda_1 [\sum_n p_n - 1] + \Lambda_2 [\sum_n r_n - 1] + \Lambda_3 [\sum_n p_n E_n - \kappa T] = \min. \quad (4)$$

Differentiating this $\partial/\partial p_n$ and equating it to zero gives as the condition for the constrained minimum

$$1 + \ln p_n - \ln r_n + \Lambda_1 + \Lambda_3 E_n = 0. \quad (5)$$

Solving Eq (5)

$$p_n = r_n \exp[-1 - \Lambda_1 - \Lambda_3 E_n] \quad (6)$$

On this basis, for a given point n , the maximum probable local reaction rate $p_n \equiv p(x_n)$ is proportional to the neighboring (at positions $x_n \pm \Delta x/2$) densities r_n of the enzyme. This makes sense since each enzyme is assumed to locally enhance the reaction, e.g. by strong hydrogen bonding, and this enhancement becomes stronger the geometrically closer the reactant is to the enzyme.

The rate p_n of reaction in Eq (6) also falls off with the local molecular bonding energy E_n . This also makes sense since the stronger the bond is the less probable it is that the molecule breaks up and contributes to the desired reactant.

Derivation for Multi-dimensional Geometry. For optimum key-lock fit, the two molecules must not only be optimally close but also each have a correct orientation. The approach to this problem requires a generalization to the use three-dimensional variables $\mathbf{x}_n \equiv (x, y, z)_n$. Here $p_n = p(x, y, z)_n$, etc. for r_n and with $\Delta x \rightarrow \Delta \mathbf{x} = (\Delta x, \Delta y, \Delta z)_n$. Also, the Kullback-Leibler distance is of the same form Eq (1a) as before,

$$\sum_{n=1}^N p(\mathbf{x}_n) \ln \left[\frac{p(\mathbf{x}_n)}{p(\mathbf{x}_n + \Delta \mathbf{x}/2)} \right] = \min. \quad (7)$$

The identical algebra Eqs (3)–(6) follow as before, with boldface quantities replacing scalars, but with the scalar E_n remaining in Eq (6) since energy is always a scalar quantity. However, an important new interpretation arises for the effect $\Delta \mathbf{x} \rightarrow \mathbf{0}$. Acknowledging this to occur in three dimensions requires the key and lock to now approach one another while in *the same orientation*. This describes a true key-lock bond. Also, now the change of reactant path so as to reduce activation energy E_a occurs in full three-dimensional space.

Note that principle Eq (7) is much more than simply a 3D version of principle (Eq 1a and Eq 1b). Consider the 3D tissue produced by multiply-folding a long string of nucleotides. From the form of Eq (7), the more regular the folding is, i.e. the more often p a given codon occurs at neighboring points \mathbf{x}_n , the closer to 1 will be the ratios in the logarithm \ln operation in Eq (7). Therefore the smaller will be their contributions to Eq (7) after the \ln is taken. Hence the smaller will be the minimum value of H_{KL} . Tissue with such low cross-entropy has low free energy and a high state of order. This might account for the vital role played by protein folding in augmenting living systems [6]. In turn, this emphasizes that H_{KL} has direct biological significance as a measure of cellular growth, despite being merely a geometrical measure.

Deriving the Arrhenius equation. The Arrhenius equation describes the dependence of reaction rates upon temperature and is empirically-derived. No enzymes are presumed present. Or equivalently, they are *equally* present at all reaction path positions [18]. Hence, we now repeat use of the minimum Kullback-Leibler principle in the special case where all enzyme

densities obey

$$r_n = r = \text{const.} \tag{8}$$

Also, for simplicity we return to the one dimensional case of scalar coordinates x_n . Recall that we used the ‘ergodic hypothesis,’ that the statistics of E at any position x_n , equals that of E over *any one path* $x_n, n = 1, \dots, N$. On this basis, and using the last identity [Eq \(3\)](#), result [Eq \(6\)](#) is, in the special case [Eq \(8\)](#)

$$P(E_n) = K \exp[-\wedge_3 E_n], \quad K = r \exp[-1 - \wedge_1]. \tag{9}$$

We also found, at [Eq \(3\)](#), that the average $\langle E \rangle = kT$. Using this in [Eq \(9\)](#) gives $\wedge_3 = K = 1/kT$. Then

$$P(E_n) = (\kappa T)^{-1} \exp[-E_n/\kappa T], \tag{10}$$

the Boltzmann energy distribution law.

At this point it is assumed that if the energy $E_n \geq E_a$, a so-called ‘activation’ level of the energy, the reaction occurs at the position x_n . But we also assumed ergodicity to hold. Therefore, the reaction occurs as often as event $E_n \geq E_a$ occurs for any one n . This shows that for any fixed energy density function $p(E_n)$ the smaller E_a is the more energy events E_n occur or, equivalently, the higher is the reaction rate.

Also, ergodicity allows us to now drop subscript n in [Eq \(10\)](#). Then using [Eq \(10\)](#) for $p(E)$ gives

$$P(E \geq E_a) = \int_{E_a}^{\infty} dE P(E) = (\kappa T)^{-1} \int_{E_a}^{\infty} dE \exp(-E/\kappa T) = \exp\left(-\frac{E_a}{\kappa T}\right). \tag{11}$$

Since each energy value E satisfying [Eq \(11\)](#) gives rise to a reaction product, this shows that the reaction rate grows as the activation energy E_a decreases.

But the analysis has ignored the fact that the molecules of the reacting medium may have a known *prior* probability A of being in the proper orientation to react. This probability should multiply result [Eq \(11\)](#).

The result is that the net probability density, or reaction rate, obeys

$$k = AP(E \geq E_a) = A \exp\left(-\frac{E_a}{\kappa T}\right), \tag{12}$$

The Arrhenius equation.

As we discussed, the optimum choice of enzyme path r_n for accomplishing the desired reaction can occur along an altered reaction path x_n requiring a *lower* activation energy E_a . This is shown by [Eq \(12\)](#) in two ways: First, the required energy values E can be smaller; and second, the resulting reaction rate k is higher. That E_a is, in fact, a minimum is shown in the Appendix *to follow* from the H_{KL} principle ([Eq 1a](#) and [Eq 1b](#)). Thus, the H_{KL} principle derives both the well-known rate effect [Eq \(12\)](#) and the fact that activation energy E_a tends to be a minimum value.

Optimization of reactant path by quantum entanglement

In the preceding, only densities p_n were optimized for a fixed enzyme density path r_n . However, further optimization can be made whereby the r_n themselves are allowed to change so as to further improve the key/lock fit. This is called ‘conformer selection’ or ‘induced fit.’[\[18\]](#). We propose two effects that potentially accomplishing this.

As noted above, enzyme function requires a tight geometric fit in which the atoms of the amino acids in to protein and the substrate molecules are separated by distances that are minimized. Suppose, as we found, their spacings $\Delta x/2$ are on the order of angstroms. At such molecular distances, quantum effects can enter in, e.g. in the form of quantum *entanglement*. This is even for semiclassical quantum effects [19]

Other authors [20], in fact, define the degree of *global entanglement* between two systems p_n, r_n as the very value of $H_{KL}(p||r)$ for the p_n, r_n obeying KL principle (Eq 1a and Eq 1b). That is: *The level of entanglement is defined by the minimized value of the Kullback-Leibler entropy, which was our very criterion (Eq 1a and Eq 1b) for the choice of the p_n .*

This also makes intuitive sense: By Eq (1a) ‘distance’ measure $H_{KL}(p||r)$ is mathematically at its absolute minimum value, of zero, when all $p_n = r_n$. This describes perfect entanglement between the the two systems p_n, r_n so that interchanges

$$p_n \leftrightarrow r_n \tag{13}$$

of the roles played by enzymes and reactants repeatedly take place. By the same token, finite values, instead, of $H_{KL}(p||r)$ allow only certain pairs of the p_n, r_n to effectively interchange roles. It results, then, that over a number of such reactions the initial molecular reactant paths $p(x_n), r(x_n), n = 1, \dots, N$ can progressively wander off to totally different ones which further upgrade the key/lock fit. These are also, in fact, *energetically preferred* since, by Eq (12), the progressively lowered threshold energy E_a is more readily provided at each such entanglement.

Discussion

Here we investigate a mechanism by which living systems use information to maintain a low entropy state far from thermodynamic equilibrium. We propose that the information encoded in the inherited sequence of nucleotides in DNA is manifested geometrically in the 3 dimensional shape of an enzyme determined by the lowest free energy state of the amino acid sequence specified by the corresponding gene. However, we note that the 3 dimensional shape of the enzyme can be extensively altered by post-translation modified. Thus, the geometry of the enzyme represents a summation of heritable information represented by its amino acid sequence and temporally variable information regarding the state of the cell and its environment which govern post translation modification.

Most simply, the information within the 3 geometry of the protein is manifested thermodynamically by the reduction in the activation energy (E_a) of the reaction catalyzed by the enzyme.

The mechanism by which information reduces the activation energy is geometric as, like a “lock and key”, the shape of the enzyme precisely fits the shape of a substrate. We investigate these spatial interactions using the Kullback-Leibler distance, Eq (1a), which is a generalization of the Shannon mutual information. In fact in many textbooks the latter is derived as a special case of the former. We demonstrate that the information of the enzyme “lock” vis a vis the shape of the substrate “key” is the equivalent of the K-L distance. Maximum information corresponds to a minimal K-L distance and, thus, the largest possible decrease in the E_a .

The observable effect of the enzyme-induced decrease in E_a is an increase in the reaction rate k , often by several orders of magnitude. This is quantified by the empirically-derived Arrhenius equation. Here we demonstrate that the Arrhenius equation can be derived from a first principle that requires minimum Kullback-Leibler divergence, (Eq 1a and Eq 1b), between a *fixed* enzyme density function and an unknown reactant function.

Here we also investigate the more recently proposed “induced fit” model in which the enzyme geometry *changes* in response to the substrate thus further improving the geometric

match. Interestingly, we find that the induced fit dynamic will occur over very small molecular distances Δx , which will potentially permit quantum entanglement effects. In particular, we find for small Δx the minimized KL entropy becomes proportional to the degree of quantum entanglement of path functions p_n, r_n . This extends prior studies suggesting quantum effects in proteins including enzymes [21–23].

Our investigation also provides general insights into the dynamics of biological information. Although it is clear that information must play a central role in the growth of living systems, the general principles that govern translation of information into biological order and function are not well defined [24]. We note that an enzyme can alter the living system in ways similar to the classic Maxwell's demon *gedanken* [25,26]. For example, a protein within a membrane can use its information (expressed as its 3 dimensional shape) to select and bind a substrate on one side of the membrane and move it into the adjacent cellular compartment [27] thus creating a spatial concentration gradient similar to the classic thought experiment [28]. However, unlike the iconic demon, enzymes can also generate a gradient *over time* [29]. That is, by greatly accelerating the rate of reaction, the concentration of substrate and products over time will be larger and smaller respectively when an enzyme is present compared to a system in which the information content of the enzyme is absent.

Finally, we note that biological information in our study is highly contextual. This is apparent, in Figs 1 and 2, as an enzyme-dependent quantitative change in activation energy E_a is dependent on both the properties of the enzyme and the properties of the substrate. Thus, in Fig 2, addition of an enzyme that is specific to the AB reaction, but not the AC reaction, lowers E_a for the AB reaction relative to that for the AC. As a result the energy E of system AB will much more often obey $E \leq E_a$ and, hence, occur much more often than the reaction AC. The information in the enzyme can, thus, be viewed as “kinetic” in reaction AB and only “potential” in the absence of the substrate. Restating this quantitatively, the information of an enzyme is defined by the KL divergence between the enzyme and a potential reactant. Further, the level of this information in each biological enzyme is converted to a thermodynamic property by the change in E_a that it evokes. Thus, the information may be either ‘potential’ or ‘kinetic,’ depending on context. The kinetic information represents the increased probability of a reaction and decreased E_a , when substrate to which it can bind is present according to principle (Eq 1a and Eq 1b). By contrast, the same enzyme but in the presence of substrate with which it *cannot react* (or in the absence of substrate) carries only potential information. It is interesting that such contextual dependence is lacking in, e.g., the pure Shannon entropy [30] measure $H_S = -\int p(x) \ln p(x)$. The algebraic difference is that the KL information is of p in the presence of context r whereas the Shannon H_S is in p by itself, *in the absence of* any context r . In summary, it is the contextual dependence of the KL information that provides its biological significance and gives rise to its function.

Appendix

The H_{KL} principle implies as well that activation energy $E_a = \text{minimum}$

Enzymes perform the critical task of lowering the activation energies E_a of chemical reactions inside the cell. For example, it is obvious from the form of the rate Eq (12) that the reaction rate k is maximized when energy E_a is a minimum value. But specifically *what effect* lowers E_a ? Could it, e.g., be our working principle Eq (1b) $H_{KL}(p||r) = \text{minimum}$? If so, the principle would now have doubled value. This is verified next.

It is convenient to work with the continuous version of H_{KL} Eq (1a), where the general coordinates x_n go over into continuous energy values $E_n = E$. Also use the route (setting $r_n = r =$

const.) Eq (8) to the Arrhenius equation. Then the principle Eq (1b) becomes

$$H_{KL} = \int dEP(E) \ln\left(\frac{-P(E)}{r}\right) = \min. \tag{A1}$$

where by Eq (10)

$$P(E) = (\kappa T)^{-1} \exp\left(-\frac{E}{\kappa T}\right). \tag{A2}$$

Expanding the \ln in principle Eq (A1) gives directly

$$H_{KL} = \int dEP(E) \ln P(E) - \ln(r) \int dEP(E) = \min. \tag{A3}$$

Using expression Eq (A2) for $P(E)$ in Eq (A3), and the normalization of $P(E)$, give

$$H_{KL} = \int_{E_a}^{\infty} dE \exp\left(-\frac{E}{\kappa T}\right) \left[-\left(\frac{E}{\kappa T}\right) - \ln(\kappa T)\right] - (\kappa T) \ln(r) = \min. \tag{A4}$$

Why is E_a the lower integration limit? Since our aim centers on the value of rate k we only integrate over those values of E that can contribute to k , and by Eq (12) this is the value E_a .

Dividing through Eq (A4) by κT and doing the integrations gives a condition

$$\frac{H_{KL}}{\kappa T} \equiv y = \exp\left(-\frac{E_a}{\kappa T}\right) \left(\frac{E_a}{\kappa T} + 1\right) - \ln(\kappa T) - \ln(r) \equiv \min. \tag{A5}$$

To attain the required minimum in H_{KL} through choice of E_a requires setting $\frac{\partial y}{\partial E_a} = 0$. Differentiating Eq (A5) in this way gives a requirement

$$\frac{E_a}{\kappa T} \exp\left(-\frac{E_a}{\kappa T}\right) = 0. \tag{A6}$$

This is accomplished by either $E_a = 0$ or $E_a = \infty$. From the result Eq (12) for the reaction rate k it is obvious that these activation energy values respectively maximize, or minimize, the rate k . Of course the case $E_a = 0$ is preferred on the basis of maximum reaction rate. However, our aim here is to show that this activation energy also follows from our overall principle (Eq 1a and Eq 1b) that $H_{KL} = \min$. Since Eq (A5) gives H_{KL} (proportional to y) we can use it to judge if the usual requirement for attaining a minimum is satisfied, namely that the second derivative $\partial^2 y / \partial E_a^2 > 0$. Taking this second derivative gives the anticipated result

$$\frac{\partial^2 y}{\partial E_a^2} = \frac{1}{(\kappa T)^2} > 0. \tag{A7}$$

Hence the case $E_a = 0$ both maximizes the reaction rate k and minimizes H_{KL} as required. By Eq (A5) zero activation energy gives a minimum H_{KL} of value

$$H_{KL} = \kappa T(1 - \ln(\kappa r T)). \tag{A8}$$

Of course attaining activation energy $E_a = 0$ is not a usual case, but the analysis shows that the closer the system gets to achieving it the higher the reaction rate is, and the smaller the KL distance is between enzyme and substrate, i.e. the better does the key fit into the lock.

Author Contributions

Conceived and designed the experiments: RG BRF. Performed the experiments: RG BRF. Analyzed the data: RG BRF. Wrote the paper: RG BRF.

References

- Gatenby RA, Frieden BR. Information theory in living systems, methods, applications, and challenges. *Bull Math Biol.* 2007; 69(2):635–57. doi: [10.1007/s11538-006-9141-5](https://doi.org/10.1007/s11538-006-9141-5) PMID: [17083004](https://pubmed.ncbi.nlm.nih.gov/17083004/)
- Frieden BR, Gatenby RA. Order in a multidimensional system. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2011; 84(1 Pt 1):011128. PMID: [21867134](https://pubmed.ncbi.nlm.nih.gov/21867134/)
- Frieden BR, Gatenby RA. Information dynamics in living systems: prokaryotes, eukaryotes, and cancer. *PLoS One.* 2011; 6(7):e22085. doi: [10.1371/journal.pone.0022085](https://doi.org/10.1371/journal.pone.0022085) PMID: [21818295](https://pubmed.ncbi.nlm.nih.gov/21818295/)
- Frieden BR, Gatenby RA. Principle of maximum Fisher information from Hardy's axioms applied to statistical systems. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2013; 88(4):042144. doi: [10.1103/PhysRevE.88.042144](https://doi.org/10.1103/PhysRevE.88.042144) PMID: [24229152](https://pubmed.ncbi.nlm.nih.gov/24229152/)
- Frieden BR, Gatenby RA. Cell development obeys maximum Fisher information. *Front Biosci (Elite Ed).* 2013; 5:1017–32.
- Albert B J A, Lewis J, Morgan D, Raff M, Roberts K, Walter P. *Molecular Biology of the Cell*: Garland Science; 2014.
- Prabakaran S, Lippens G, Steen H, Gunawardena J. Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip Rev Syst Biol Med.* 2012; 4(6):565–83. doi: [10.1002/wsbm.1185](https://doi.org/10.1002/wsbm.1185) PMID: [22899623](https://pubmed.ncbi.nlm.nih.gov/22899623/)
- C GM.. *The Cell: A Molecular Approach*. 2nd edition. Sunderland, MA: Sinauer Associates; 2000.
- Ghirlanda G. Computational biochemistry: old enzymes, new tricks. *Nature.* 2008; 453(7192):164–6. doi: [10.1038/453164a](https://doi.org/10.1038/453164a) PMID: [18464727](https://pubmed.ncbi.nlm.nih.gov/18464727/)
- Koshland DE. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci U S A.* 1958; 44(2):98–104. PMID: [16590179](https://pubmed.ncbi.nlm.nih.gov/16590179/)
- Koshland DE Jr. The active site and enzyme action. *Adv Enzymol Relat Subj Biochem.* 1960; 22:45–97. PMID: [13753331](https://pubmed.ncbi.nlm.nih.gov/13753331/)
- Thoma JA, Koshland DE Jr. Stereochemistry of enzyme, substrate, and products during beta-amylase action. *J Biol Chem.* 1960; 235:2511–7. PMID: [13837904](https://pubmed.ncbi.nlm.nih.gov/13837904/)
- Alberts B, Bray D, Hopkin K, Johnson A, Lewis J, Raff M, et al. *Protein Structure and Function*. In: *Essential Cell Biology*. 3rd ed. New York: Garland Science, Taylor and Francis Group, LLC; 2010. p. 120–70.
- Kullback S, Leibler RA. On Information and Sufficiency. 1951:79–86. doi: [10.1214/aoms/1177729694](https://doi.org/10.1214/aoms/1177729694)
- K S.. *Information Theory and Statistics*. New York: Wiley; 1959.
- Thoma JA, Koshland D.E. Competitive Inhibition by Substrate during Enzyme Action. Evidence for the Induced-fit Theor. *Journal of the American Chemical Society.* 1960; 82(13):3329–33.
- Koshland DE. The Key-Lock Theory and the Induced Fit Theory. *Angew Chem Int Ed Engl.* 1994; 33:2735–378.
- Peacock DG, Richardson J.F. *Chemical Engineering*. 3rd ed. Oxford: Elsevier Ltd; 1994.
- Jacquod P. Semiclassical time evolution of the reduced density matrix and dynamically assisted generation of entanglement for bipartite quantum systems. *Phys Rev Lett.* 2004; 92(15):150403. doi: [10.1103/PhysRevLett.92.150403](https://doi.org/10.1103/PhysRevLett.92.150403) PMID: [15169271](https://pubmed.ncbi.nlm.nih.gov/15169271/)
- Sarovar M, Ishizaki A., Fleming G.R., Whaley B. Quantum entanglement in photosynthetic light-harvesting complexes. *Nature Physics.* 2010; 6:462–7. doi: [10.1038/nphys1652](https://doi.org/10.1038/nphys1652)
- Gray HB, Winkler JR. Electron tunneling through proteins. *Q Rev Biophys.* 2003; 36(3):341–72. doi: [10.1017/S0033583503003913](https://doi.org/10.1017/S0033583503003913) PMID: [15029828](https://pubmed.ncbi.nlm.nih.gov/15029828/)
- Nagel ZD, Klinman JP. Update 1 of: Tunneling and Dynamics in Enzymatic Hydride Transfer. *Chem Rev.* 2010; 110(12):Pr41–Pr67. doi: [10.1021/cr1001035](https://doi.org/10.1021/cr1001035) PMID: [21141912](https://pubmed.ncbi.nlm.nih.gov/21141912/)
- Nagel ZD, Klinman JP. Tunneling and dynamics in enzymatic hydride transfer. *Chem Rev.* 2006; 106(8):3095–118. doi: [10.1021/cr050301x](https://doi.org/10.1021/cr050301x) PMID: [16895320](https://pubmed.ncbi.nlm.nih.gov/16895320/)
- Mandal D, Jarzynski C. Work and information processing in a solvable model of Maxwell's demon. *Proc Natl Acad Sci USA.* 2012; 109(29):11641–5. doi: [10.1073/pnas.1204263109](https://doi.org/10.1073/pnas.1204263109) PMID: [22753515](https://pubmed.ncbi.nlm.nih.gov/22753515/)
- Maruyama K, Nori F, Vedral V. Colloquium: The physics of Maxwell's demon and information. *Rev Mod Phys.* 2009; 81(1):1–23. doi: [10.1103/RevModPhys.81.1](https://doi.org/10.1103/RevModPhys.81.1)

26. Shenker OR. Maxwell's Demon 2: Entropy, classical and quantum information, computing. *Stud Hist Philos M P*. 2004; 35b(3):537–40.
27. Saier MH Jr, Tran CV, Barabote RD. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res*. 2006; 34(Database issue):D181–6. doi: [10.1093/nar/gkj001](https://doi.org/10.1093/nar/gkj001) PMID: [16381841](https://pubmed.ncbi.nlm.nih.gov/16381841/)
28. Fernandez-Belda F, Inesi G. Transmembrane gradient and ligand-induced mechanisms of adenosine 5'-triphosphate synthesis by sarcoplasmic reticulum adenosinetriphosphatase. *Biochemistry-U.S.* 1986; 25(24):8083–9.
29. Lin F, Saadi W, Rhee SW, Wang SJ, Mittal S, Jeon NL. Generation of dynamic temporal and spatial concentration gradients using microfluidic devices. *Lab Chip*. 2004; 4(3):164–7. doi: [10.1039/b313600k](https://doi.org/10.1039/b313600k) PMID: [15159771](https://pubmed.ncbi.nlm.nih.gov/15159771/)
30. Strait BJ, Dewey T.G. The Shannon information entropy of protein sequences. *J Biophys*. 1996; 71:148–55.