Published online (day) (month) 2010 *Nucleic Acids Research, 2010, Vol. ?, No. ?* ***1–14***
doi:10.1093/nar/gkn000

70% efficiency of bistate molecular machines explained by information theory, high dimensional geometry and evolutionary convergence

Thomas D. Schneider[1]*

[1]Center for Cancer Research Nanobiology Program, National Cancer Institute, Building 469, Room 215, P.O. Box B, Frederick, MD USA 21702-1201

Received January 19, 2010; Revised April 8, 2010; Accepted April 29, 2010

ABSTRACT

The relationship between information and energy is key to understanding biological systems. We can display the information in DNA sequences specifically bound by proteins by using sequence logos, and we can measure the corresponding binding energy. These can be compared by noting that one of the forms of the second law of thermodynamics defines the minimum energy dissipation required to gain one bit of information. Under the isothermal conditions that molecular machines function this is $\mathcal{E}_{min} = k_B T \ln 2$ joules per bit (k_B is Boltzmann's constant and T is the absolute temperature). Then an efficiency of binding can be computed by dividing the information in a logo by the free energy of binding after it has been converted to bits. The isothermal efficiencies of not only genetic control systems, but also visual pigments are near 70%. From information and coding theory, the theoretical efficiency limit for bistate molecular machines is $\ln 2 = 0.6931$. Evolutionary convergence to maximum efficiency is limited by the constraint that molecular states must be distinct from each other. The result indicates that natural molecular machines operate close to their information processing maximum (the channel capacity), and implies that nanotechnology can attain this goal.

INTRODUCTION

Measuring information and energy in biological systems

To address the relationship between information and energy in biological systems requires first being able to measure each one. Standard methods for measuring the energy dissipation of molecular interactions are well established (1, 2), but the corresponding measure of information (3) is rarely determined. To make a measure of information that can be compared to an energy difference, that measure must express a state change corresponding to the binding interaction. The easiest systems to work with are DNA binding proteins since the patterns to which they bind can be readily determined by sequencing technologies, and from these data one can compute the information gained in the process (3). The state change that can be measured for both information and energy is between the molecule being anywhere on the DNA (but already non-specifically bound to the DNA, the *before* state), and molecules bound to specific functional sites (the *after* state). To make a comparison, not only must the state changes be the same but also the number of molecules involved must be equivalent. In this paper measurements for both energy and information are reported on a per-molecule basis.

Information is displayed by sequence logos

For the information measure, sequence logos are a widely used graphical representation of aligned biological sequences such as DNA or RNA binding sites or protein motifs (4, 5). In a conventional logo, the symbols of the polymer alphabet are stacked one on

*To whom correspondence should be addressed. Tel: 301-846-5581; Fax: 301-846-5598; Email: toms@alum.mit.edu. version = 2.93 of emmgeo.tex 2010 May 19

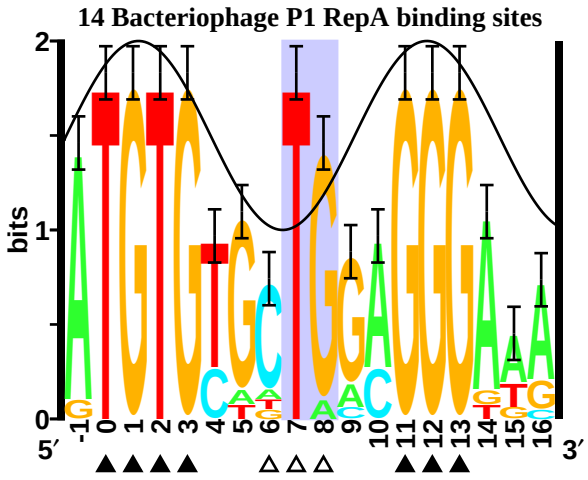

Figure 1. Sequence logo for RepA binding sites from bacteriophage P1 (GenBank Accession K02380.1). The height of each letter is proportional to the corresponding base frequency at that position in the binding sites. The height of the entire stack is the sequence conservation in bits (with error from the small sample size shown) (12, 13, 14). This often varies according to a sine wave. Where the DNA faces the binding protein the information can be up to 2 bits (fully conserved, closed triangles), while in the minor groove not more than 1 bit can be conserved in B-form DNA (open triangles). The variation between these two because of the twist of the DNA helix is shown by the sine wave (7, 8). Comparing the logo to the sine wave, anomalous positions having more than 1 bit in the minor groove are revealed to be in positions +7 and +8 (blue box), suggesting that the sites are not B-form DNA when bound by RepA.

top of another, with their heights made proportional to their frequency at that position. The symbols are sorted by frequency so that the most common letter is on top. The utility of logos is that the sequence conservation is indicated by scaling the entire stack of letters to represent the information at that position in the binding site. For example, because of the structure of DNA, protein contacts into the major groove can be fully conserved at 2 bits of information but the minor groove has half the maximum sequence conservation (6) and so cannot exceed 1 bit of information. Many logos show this effect (7, 8). However, the sequence logo for bacteriophage P1 RepA binding sites (Fig. 1) has, between two well conserved major groove contact regions (0 to +3 and +11 to +13), a striking anomaly in the minor groove at +7 where conservation is near 2 bits (7, 9). Similar anomalies are observed in other proteins that bind DNA replication origins (8, 10) and these imply that the DNA is not B-form. Indeed, further experimental work revealed that after binding to a DNA replication origin, RepA probably flips the conserved T out of the helix to initiate DNA replication (11). In information theory, as with any other well-established theory, anomalies can lead to new biological understanding.

The vertical scale of a sequence logo is given in bits of information. A bit is the amount of information needed to choose between two equally likely possibilities. In the case of nucleic acids, there are four possible bases. These can be arranged into two sets, for example the purines (A and G) and the pyrimidines (C and T). One bit of information is sufficient to choose between the purines and pyrimidines and a second bit of information distinguishes the exact base. Thus sequence logos for binding sites have a scale from 0 to 2 bits.

Comparing information to binding energy

Sequence logos provide a precise, quantitative measurement of the information in binding sites. How is this related to the binding energy? Previous work (15, 16) assumed that the binding energy determines the sequence conservation in a one-to-one function so that for each binding site there would be only one energy that is proportional to one number of bits. However, a protein could evolve to bind to the same sequence with either more or less energy, just as a coin flipped to different heights still supplies no more than 1 bit of information, so the relationship between information and energy is an inequality. The lower bound of energy dissipation can be determined from the second law of thermodynamics to be $k_B T \ln 2$ joules per bit where k_B is Boltzmann's constant and T is the absolute temperature (17, 18, 19). In this paper we recognize that this second law relationship can be used as an ideal conversion factor to express binding energy dissipation as the maximum number of bits that could be gained. By comparing this potential maximum bits to the actual information observed in a logo, we form an efficiency. However, it is important to clarify the relationship between this new thermodynamic definition of efficiency and the previously defined thermodynamic efficiency.

Classical thermodynamic efficiency

In an automobile, burning fuel expands to drive the engine. Because it operates between two temperatures T_{hot} and T_{cold} such a heat engine has the classical Carnot efficiency of

$$\eta_{\text{Carnot}} = \frac{T_{hot} - T_{cold}}{T_{hot}} \qquad (1)$$

(20, 21). Jaynes noted that when one uses the Carnot formula for a biological system having 70% efficiency, one gets an anomalous result: at the temperature of a warm day, T_{cold} = 300 K, and equation (**1**) gives T_{hot} = 1000 K, which would burn tissue (22). This absurd result indicates that this thermodynamic formula does

not apply to most biological systems since molecules inside cells function at a single temperature (22, 23). For example, in the retina of the eye the protein rhodopsin detects light (24). Thermal equilibrium is attained within picoseconds after rhodopsin absorbs a photon (25, 26). Likewise, a DNA binding protein such as the restriction enzyme EcoRI, when bound nonspecifically to DNA, rapidly comes to local equilibrium with the surrounding solution (21). Once EcoRI has moved by Brownian motion to its specific binding sequence, 5′ GAATTC 3′, it binds and releases heat. The heat dissipates, leaving the DNA and protein again at local equilibrium. Since the final temperature is the same as the initial temperature, the Carnot efficiency is zero. Hence it cannot be used in molecular biology.

Communications efficiency

Yet a precisely defined, practical measure of efficiency is essential to characterize and understand biological processes. In this paper we show how an efficiency derived from both the second law of thermodynamics and information theory can be applied to isothermal biological processes. The key pieces of information theory needed to do this were published by Claude Shannon in 1948 (12) and 1949 (27). Then, in 1959 Pierce and Cutler used information theory to define an efficiency measure for satellite communications (28, 29),

$$\epsilon = \frac{\ln\left(\frac{P}{N}+1\right)}{\frac{P}{N}} \qquad (\mathbf{2})$$

where P/N is the 'signal-to-noise ratio', the power P dissipated at the receiver in joules per second, versus the thermal noise power N interfering with the signal there. This formula was derived from Claude Shannon's famous channel capacity equation,

$$C = W\log_2\left(\frac{P}{N}+1\right) \quad \text{(bits per second)} \qquad (\mathbf{3})$$

in which the bandwidth W defines the range of frequencies used in the communications as, for example, by a radio station (12, 13, 27). The channel capacity theorem states that as long as the data rate R (also in bits per second) is less than or equal to the channel capacity C, communication can be established with as few errors as desired. To reach this ideal requires that the messages be coded to protect them against noise. For example, Morse code can replace verbal communications in noisy situations. Likewise, the 8th bit of an ASCII computer character (byte) (30), which is known as a 'parity bit', can be set so that the total number of 1's is even. If an odd number of 1's is received, an error is detected. Sixty years of developing sophisticated codes and computer chips to implement them has led to reliable modern communications, including cell phones, the internet and interplanetary data transmissions. In this paper we demonstrate the application of information theory to an equally broad range of molecular machines.

Molecular machine capacity

We have previously shown that a formula equivalent to the channel capacity, equation (**3**), can be developed for molecular machine states:

$$C_y = d_{space}\log_2\left(\frac{P_y}{N_y}+1\right) \quad \text{(bits per operation)} \quad (\mathbf{4})$$

where d_{space} is the number of independent parts of the molecular machine (23), P_y is the energy dissipation from the machine per operation and N_y is the thermal noise power interfering with the machine during an operation. The subscript y indicates that the coding space is for mechanical potentials instead of voltage potentials. That is, the model is for a physical object such as a weight on a spring instead of an electrical oscillator built from capacitors and inductors (31).

The units in equation (**4**) are 'bits per operation', in which an 'operation' is, for example, moving from nonspecific to specific DNA binding by EcoRI so operations in equation (**4**) replace seconds in equation (**3**). Both capacity equations only apply to living things because the key concept used to derive them is that messages and molecular states can be distinct (32). This additional constraint does not derive from physics or thermodynamics; having discrete molecular states is a biological criterion imposed by natural selection.

Rhodopsin, for example, has two biologically important physical states: not having seen a photon and having seen one. If these states were not stable and distinct, the molecule would rapidly switch between them because of thermal impacts, giving an animal the impression that there is light when in the dark. These animals will be eliminated by natural selection, leaving only those who have evolved sufficiently distinct states. Likewise, if EcoRI in the bacterium *Escherichia coli* were to bind to incorrect positions on the DNA other than GAATTC, the genomic DNA would be destroyed because only that sequence is protected from EcoRI digestion by the corresponding DNA methylase (33). The extreme precision of restriction enzymes (34, 35, 36, 37, 38) and the thermal stability of rhodopsin (39) are well known but the underlying fundamental reason has not been widely appreciated. Shannon's channel capacity theorem, as applied to molecules (23), guarantees that by appropriate coding it is possible for a molecular machine to evolve distinct states, and once it has done so, it can operate at its capacity with as few errors as is necessary for

survival. It is important to note that the channel capacity is an ideal upper bound that cannot be exceeded because thermal noise cannot be avoided by molecules. However, unless there are additional constraints, we might anticipate that biological systems can evolve to this limit.

RESULTS AND DISCUSSION

Molecular machine isothermal efficiency

Since the capacity equation and theorem can be extended from communications systems to states of molecules (23), the efficiency can also be extended, and the resulting formula is equivalent to equation (**2**). The derivation is as follows. Consider a coin. Neglecting the unstable condition of balancing on the edge, a coin can have two states, heads up and tails up. When a coin has kinetic energy it rapidly switches between these states as when, for example, it bounces around in a box. For the coin to settle down to one state or the other, it must dissipate energy to the surroundings. The minimum energy dissipation, derived from the channel capacity or the second law of thermodynamics (with the constraint that the temperature is constant) (17, 18), is

$$\mathcal{E}_{min} = k_{\mathrm{B}} T \ln 2 \quad \text{(joules per bit)} \qquad (\mathbf{5})$$

where k_{B} is Boltzmann's constant (joules per kelvin), T is the absolute temperature (kelvin) and $\ln 2$ gives the units of 'per bit'. Obviously a coin will dissipate much more energy than this minimum because it is an inefficient macroscopic device. How much more can be defined by the relationship between the dissipated energy P_y and the information C_y:

$$\mathcal{E} \equiv \frac{P_y}{C_y} \quad \text{(joules per bit)}. \qquad (\mathbf{6})$$

In the limit as $P_y \to 0$, $\mathcal{E} \to \mathcal{E}_{min}$, so $\mathcal{E} \geq \mathcal{E}_{min}$ (18).

The efficiency of the coin or molecular machine is then defined as the minimum possible energy dissipation divided by the actual dissipation (using $N_y = d_{space} k_{\mathrm{B}} T$ (18)):

$$\epsilon_t \equiv \frac{\mathcal{E}_{min}}{\mathcal{E}} = \frac{\ln\left(\frac{P_y}{N_y}+1\right)}{\frac{P_y}{N_y}} \quad \frac{\text{(joules per bit)}}{\text{(joules per bit)}}. \qquad (\mathbf{7})$$

Notably, the isothermal efficiency is exclusively a function of the power-to-noise ratio, P_y/N_y. Using the channel capacity theorem, it can be shown that the efficiency of a real measurable system, ϵ_r, cannot exceed the theoretical limit defined by ϵ_t, as shown in Fig. 2. Both the Carnot efficiency [equation (**1**)] and the efficiency developed here [equation (**7**)] are derived using the second law of thermodynamics (22), but only the latter applies to isothermal processes.

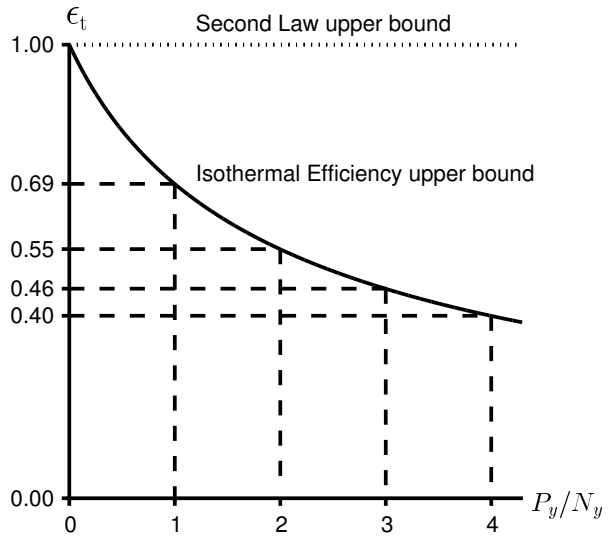

Figure 2. Efficiency curve for isothermal molecular machines. $\epsilon_t = \frac{\ln\left(\frac{P_y}{N_y}+1\right)}{\frac{P_y}{N_y}}$.

Computations of the isothermal efficiency

We now show how to apply this theory to biological processes, using EcoRI as an example. This molecule precisely selects GAATTC from all possible hexamers on DNA. To choose a single base, such as the first G, requires 2 bits (3, 13). For example, one may ask "Is it a purine (A or G)?" (a single bit will answer this question) and "Is it in the set A or T?" (a second one-bit answer). Because bits are additive (12) the total information needed to specify GAATTC is $6 \times 2 = 12$ bits and EcoRI 'gains' 12 bits when it binds by reducing its positional entropy along the DNA string by that amount (3, 40). This can be displayed graphically with a sequence logo, as shown in Fig. 3.

Binding, however, requires that the molecules stick together and to do so some energy must be dissipated (46), as in the example of the bouncing coin. This energy dissipation can be measured by electromobility shift assays (1) or directly by microcalorimetry (2), and it is expressed as the specific binding constant K_{spec}, the ratio of specific binding at GAATTC (K_s for the *after* state) to nonspecific binding anywhere on the DNA (K_n for the

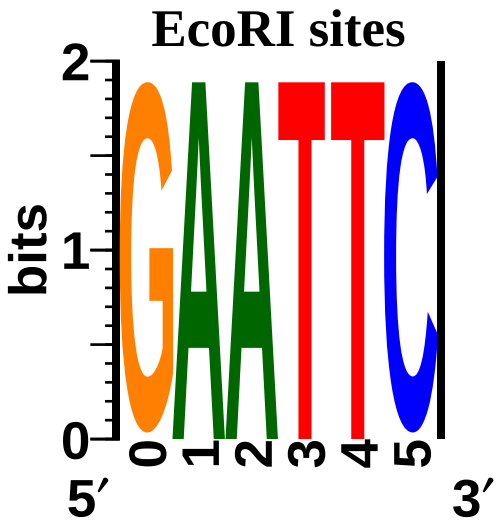

Figure 3. Sequence logo (4) of the 5′ GAATTC 3′ sequences bound by the restriction enzyme EcoRI (41). As can be seen by summing the information of 2 bits over 6 positions, the total information is 12 bits per site. In most binding sites on DNA the information varies with position in the site (Fig. 1) (3, 7, 42), but since information is additive when positions are independent (12), the total information can also be computed in those cases (43). If the positions are not independent, the correlations can be accounted for by the appropriate computation (44, 45). In this case since there is no variation in the EcoRI site, the positions are independent and the total 12 bits is obtained by a simple sum (12).

before state):

$$K_{spec} = K_s / K_n \qquad (8)$$

For EcoRI, K_s and K_n have been measured, and K_{spec} is $1.6 \times 10^5 \pm 1.4 \times 10^4$ on the λ *sr*I 2 site (47). The specific binding energy is

$$\Delta G^\circ_{spec} = -k_B T \ln K_{spec} \quad \text{(joules per binding)} \quad (9)$$

where we have chosen k_B so as to give results on a per-molecule basis, instead of the gas constant R which gives joules per mole. ΔG°_{spec} represents the maximum energy available for the selection process (48, 49). Since $\mathcal{E}_{min}$ is the ideal minimum energy dissipation per bit, equation (**5**), we introduce the use of $\mathcal{E}_{min}$ as an ideal conversion factor to determine the maximum number of bits corresponding to a given energy dissipation:

$$R_{energy} \equiv -\Delta G^\circ / \mathcal{E}_{min} = \log_2 K_{spec} \qquad (10)$$
(bits per binding)

which is a remarkably simple equation. In the case of EcoRI, $R_{energy} = 17.3 \pm 0.1$ bits/binding. That is, the molecule could have, by the second law of thermodynamics, made an average of 17.3 discrete yes-no selections for the given energy dissipation. But from the sequence GAATTC we know that it only selects $R_{sequence} = 12$ bits per binding. This must be less than the capacity,

$$R_{sequence} \leq C_y \qquad (11)$$

so in parallel with equation (**6**) we can define

$$\mathcal{E}_r = -\Delta G^\circ / R_{sequence} \quad \text{(joules per bit)}. \qquad (12)$$

$\mathcal{E}_r$ must exceed $\mathcal{E}_{min}$ (18). The observed efficiency measures the discrepancy between the information and the energy as

$$\epsilon_r = \mathcal{E}_{min} / \mathcal{E}_r \qquad (13)$$
$$= R_{sequence} / R_{energy} \qquad (14)$$
$$= 12 / 17.3 \pm 0.1$$
$$= 69.4 \pm 0.4\%$$

by substituting equations (**12**) and (**10**) into (**13**). Other binding sites for EcoRI give similar but slightly different efficiencies: λ *sr*I 5 is $73.7 \pm 0.6\%$ and pBR is $66.7 \pm 1.4\%$, suggesting unaccounted for experimental variation or some influence of the surrounding sequence that was not used in the information measure. However, EcoRI is used here as an example because it is a well-characterized DNA binding protein that has both non-specific binding data and reported errors.

As a more general example, let's calculate the isothermal efficiency of the RepA protein binding to its DNA sites. The sum of the varying sequence conservation in Fig. 1 for the range from −1 to +16 is $R_{sequence} = 24.52 \pm 1.17$ bits/site (reporting the standard error of the mean for the individual information distribution (50)). The non-specific binding energy is not known so we will (tentatively) assume it is zero (*i.e.* $\log_2 K_n = 0$). The binding constant K_s has been reported as $K_D = 0.10(\pm 0.09)$ nM (51). Taking $\log_2$ we find $R_{energy} = \log_2 K_s - \log_2 K_n = 33.22 \pm 1.30$ (bits per site). So from equation (**14**) the efficiency is $\epsilon_r = 0.74 \pm 0.05$. If there is non-specific binding, it would lower R_{energy} and raise the efficiency.

Including EcoRI and RepA, the information used by DNA binding proteins for a variety of genetic control systems has been measured (3, 4, 52) and 18 of 19 of them also have efficiencies near 70% (manuscript in preparation). Strikingly, the quantum efficiency of rhodopsin ($66 \pm 2\%$ for 12 species) (53), bacteriorhodopsin ($67 \pm 4\%$) (54) and photoactive yellow protein (64%) (55) are also around 70%. Why are all these molecular machines ~70% efficient?

Evolution of isothermal efficiency up to 70%

Each of the molecular machines having 70% efficiency functions isothermally and all of them select a discrete state from amongst several possible states. EcoRI and other genetic recognizers select patterns on DNA (3), while the rhodopsin protein and its retinal chromophore selects the stable bathorhodopsin (metarhodopsin II) configuration which triggers reactions leading to a nerve impulse (24, 56).

As Tribus and McIrvine have shown (57), energy dissipation from human activities and machines such as computers is orders of magnitude higher than the thermodynamic limit [equation (**5**)]. The corresponding efficiencies of human activities are on the order of 10^{-21}, while computers (in 1997) had reached 10^{-6} (58). Using equation (**2**), Pierce and Cutler reported that for amplitude modulation (AM) radio signals, "good quality speech or television is a factor of several hundred times less efficient than the ideal", while for frequency modulation (FM) the efficiency can be at most 5% (28). The biological efficiencies near 70% are high by human technology standards. However, this means that for every 100 photons absorbed by rhodopsin, 30 are wasted as heat (25, 56). It would be a great evolutionary advantage to see those lost 30 photons. Any DNA-protein contacts that dissipate extra energy while not contributing information to help locate sites of EcoRI, would be lost by mutations. Because the information needed to locate binding sites is fixed (3, 40), this atrophy drives the efficiency up for nucleic acid recognizers. Something must be preventing these molecular machines from exceeding 70% efficiency.

A power to noise ratio exceeding 1 explains 70% efficiencies

The observation that many molecular machines are 70% efficient can be understood by using the isothermal efficiency given by equation (**7**). In Fig. 2, the second law corresponds to the horizontal dashed line at 100%. Shannon's channel capacity theorem can be used to demonstrate that the region between the second law bound and the isothermal efficiency curve cannot be reached by any system. The curve shows that an efficiency of 70% corresponds to a P_y/N_y ratio of 1. Because the curve defines an upper bound, an upper limit on the efficiency corresponds to a lower limit on P_y/N_y. So the efficiencies of 70% can be explained by proposing that $P_y > N_y$.

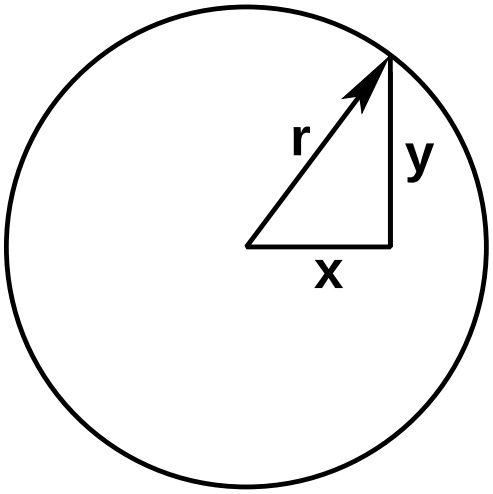

Figure 4. Geometry for combining two Gaussian distributions.

Communications coding spaces

The reason that the energy dissipation P_y barely exceeds the thermal noise N_y can be understood by considering the elegant geometrical derivation of the channel capacity by Shannon in 1949 (27), in which he represented messages as points in a high dimensional coding space. A complex message, such as a song, can be sampled and digitized to produce a stream of bits represented by voltage pulses in a wire. The amplitude of the first pulse is independent of that of the second pulse and, more importantly, the thermal noise which interferes with both affects each independently. There are four possible combinations for two pulses and these may be represented in two dimensions as a square. If we introduce a third pulse, the possible combinations are represented by the corners of a cube. A message consisting of 100 pulses is then expressed as a point on the corner of a 100 dimensional hypercube. These points representing messages can be placed into other arrangements besides cubic spacing to form different lattices (59).

Thermal noise interferes with the pulses, smearing them out to a Gaussian distribution in each dimension. The combination of several independent Gaussian distributions forms a spherical distribution (23, 27), as discussed in reference (32). To see this, we note that since the noise of two pulses is independent, we can graph the magnitude noise of the first pulse on the x axis and the second on the y axis to form a 'noise vector' in the direction to the point (x,y) as shown in Fig. 4. Because the noise is Gaussian, the probability of having a disturbance x has the form

$$P(x) = e^{-x^2} \qquad (\mathbf{15})$$

(ignoring the constants, which will drop out in a moment) and for the second pulse,

$$P(y) = e^{-y^2}. \quad (\mathbf{16})$$

The probability of being at a point (x,y) on the x-y plane is

$$P(x,y) = P(x)P(y) \quad (\mathbf{17})$$

since x and y are independent. Combining these three equations gives

$$P(x,y) = e^{-(x^2+y^2)} = e^{-r^2}. \quad (\mathbf{18})$$

Geometrically, x and y are the legs of a triangle with hypotenuse r. If the joint probability $P(x,y)$ is a constant, then by equation (**18**) r is a constant and the possible solutions trace out a circle. Furthermore, this is true for every joint probability, so the overall distribution is circularly symmetric at every radius (27, 32). This argument extends to a third pulse, so the noise is represented by a sphere around the original signal point in three dimensional space. In general, when two or more independent Gaussian distributions are combined orthogonally in one space they form a fuzzy spherical shell, and at higher dimensions the sphere is hollow and well defined because no matter which direction the thermal noise vector is pointing in the high dimensional space, its magnitude is approximately constant over many pulses (21, 23, 27, 60).

A transmitted message is received as a point somewhere on this sphere around the original message point (23, 27, 32). Given the point on the sphere, the receiver merely needs to choose the closest sphere center to remove the noise. This is possible as long as the spheres do not intersect significantly. The channel capacity formula [equation (**3**)] was derived by counting how many nonoverlapping thermal noise spheres can be packed together into the larger sphere defined by the power dissipation and the thermal noise. Coding of the messages is defined by the sphere locations in the lattice packing (27).

Molecular coding spaces

To model the coding space of molecules, we replace the voltage pulses with a mechanical equivalent; 'pins' in a lock make a good analogy (23). Each pin is a cluster of atoms of a molecule that moves as a unit independently of the motion of other pins. To the degree that the pins are not independent, the effectiveness of the lock is reduced and, correspondingly, a molecular machine will function below capacity. (Alternatively, a pin could be represented by a vibrational mode of the molecule. By definition these normal modes are independent.) As in a lock, the independently moving pins cooperate to change the state of the molecule.

Since the thermal noise impacting on the molecule is Gaussian and there are many pins, the energetic state of a molecular machine such as EcoRI or rhodopsin can also be represented as a sphere in a high dimensional coding space Y and an equivalent capacity can be derived for these molecules [equation (**4**)] (19, 23, 32). For example, when rhodopsin is in the dark, the thermal noise impacting on it from all directions can be represented by a sphere. The direction of this energy changes randomly by Brownian motion (19). As long as two thermal noise spheres do not intersect significantly, the molecule will only rarely switch between the states and there will be few errors. Upon absorbing a photon the radius of the sphere expands.

The *before*, *forward* and *degenerate* coding states

Different frequencies of light have different energies, but over most of the spectrum the efficiency of rhodopsin is constant (61). The reason for this effect is that after absorption, the excess photon energy is lost, leaving rhodopsin in a high energy metastable state (21, 54) encompassing several possible lower energy molecular states. It is from this '*before*' state that the molecule must choose a new '*forward*' configuration (23) or it will collapse back to its original state, which we coin the '*degenerate*' state. Fig. 5 shows these three state spheres.

In Fig. 5, both the *forward* and the *degenerate* spheres have radii determined by thermal noise, and both are enclosed by the *before* sphere. However, there is one unusual feature of high dimensional space that must be handled to correctly draw a diagram of the relationships between the three spheres. Because it represents the same thermal noise energy as the *degenerate* sphere, the *forward* state sphere is represented by a straight line segment having a length the same as the diameter of the *degenerate* sphere. This flattened representation, which Shannon used in his proof of the channel capacity theorem (27), can be understood by considering an analogy for noise, the effect of winds on an airplane in a turbulent storm. As the plane flies forward it is buffeted in three dimensions. Two of these throw it off course while the third advances or retards it. Likewise, if the plane were flying in a 100 dimensional space, 99% of the buffeting wind would throw it off course, while only 1% would affect its progress. Thermal noise affects molecular decisions in a similar way. Thus, with respect to the direction of motion between molecular states, thermal noise can be represented as a flat disc at 90° to that motion; as engineers (following Shannon) we can neglect the 1%. However, if 1%

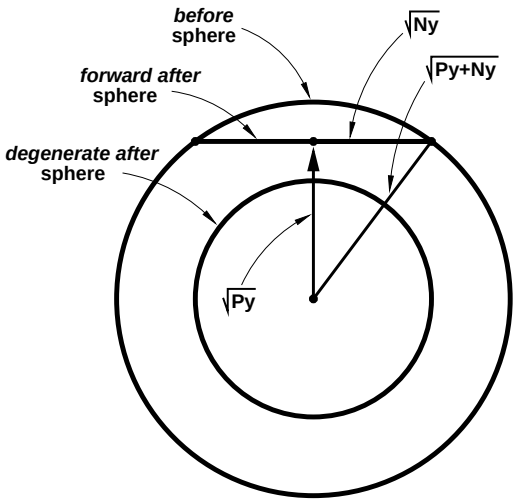

Figure 5. Velocity-potential state diagram for bistate molecular machines. See (23) for more details about this space and these spheres. Three spheres in a high dimensional space are represented as two concentric circles and a line segment. The outer circle represents the *before* sphere with radius $\sqrt{P_y + N_y}$; the inner circle represents the *degenerate after* sphere and the horizontal line segment represents the *forward after* sphere. Bistate molecular machines have these two *after* spheres, both of which have a radius determined by thermal noise, $\sqrt{N_y}$. The vertical arrow indicates the direction and magnitude of the velocity, $\sqrt{P_y}$, that the molecular machine moves at to escape from the *degenerate* sphere state. Note that the Pythagorean theorem defines the relationship between the three labeled line segments which form a triangle since the direction of motion is perpendicular to the *forward* sphere line segment (23). The diagram is derived from Fig. 5 of Shannon's 1949 paper (23, 27). For an alternative diagram and proof, see the Appendix.

leads to too much error (by state switching from the *forward* back to the *degenerate* state and *vice versa*), the errors may be reduced further by evolving a higher dimensionality. Many molecular machines are likely to operate in this realm of high dimensions because the potential dimensionality of a molecule is the number of degrees of freedom, and this depends on the number of atoms, n, according to $3n - 6$ (3 dimensions of motion for each atom, less 3 translational motions and 3 rotations of the whole molecule, as measured in infrared spectroscopy). Only some of the atoms can be involved in the recognition process required to define states, but large molecules such as EcoRI on DNA [$n = 9106$ (62)] and rhodopsin [$n = 5511$ (63)] could be operating in many dimensions. So for these examples the error could be negligible, and in Fig. 5, the *forward after* sphere is drawn as a straight line segment.

If the *forward* sphere intersected the *degenerate* sphere, then rhodopsin could switch between these states merely by thermal noise. Thus the *forward* state must be sufficiently displaced from the *degenerate* state. The *degenerate* sphere is exactly in the center of the *before* sphere because the photon excitation causes high energy vibrations of the entire molecule in no particular direction so these two spheres are drawn as concentric circles. Having absorbed a photon, rhodopsin is in the *before* state ready to 'choose' between the *degenerate* and the *forward* state.

Coding space explanation of $P_y/N_y > 1$

For a molecular machine, the time unit is defined by the operation which selects the *after* states, so power is equal to the energy dissipated during one state selection. Furthermore, the kinetic energy of each thermally vibrating pin is proportional to the square of its maximum velocity when the potential energy is zero. Combining these two ideas, we see that the maximum pin velocity is proportional to the square root of the power. So, given the available energy P_y, the maximum velocity that the molecule can attain to escape the *degenerate* sphere state is $\sqrt{P_y}$ (23). This is shown as an arrow connecting the centers of the *degenerate* and *forward* states in Fig. 5.

Likewise, the unavoidable thermal noise energy N_y that flows into and through the molecule during a molecular machine operation interferes with the corresponding power dissipation vector and has a magnitude of $\sqrt{N_y}$ (23). In the high dimensional space, most of the noise is at right angles to the power, so together these two orthogonal vectors define the radius of the *before* sphere to be $\sqrt{P_y + N_y}$. From the *before* state rhodopsin will dissipate energy and select either the *forward* or the *degenerate after* state. These two states will be distinct from each other only if they do not intersect, which means that $\sqrt{P_y}$ must exceed the radius of the *degenerate* thermal noise sphere, $\sqrt{N_y}$, and so $P_y/N_y > 1$ and the efficiency cannot be higher than $\ln(2) = 69.3\%$ by equation (**7**). Fig. 6 shows the geometrical configuration when $P_y = N_y$. This diagram and the efficiency equation explain why many molecular machines have efficiencies near 70%.

Why is the *degenerate* sphere avoided as much as possible by molecular machines? For every point on the *before* sphere there is a corresponding point on the *degenerate* sphere. They represent the same motions except that the *before* sphere motions have more energy. Yet in the *before* sphere all possible substates are available to choose from since it encloses many *after* spheres, while in the *degenerate* sphere the energy has been dissipated so there is no possibility of making choices anymore. If the molecular machine enters the *degenerate* sphere it would have wasted its energy. For rhodopsin, it would return to its original state and fail to detect a photon. Worse, if the *degenerate* sphere intersected the *forward* sphere for rhodopsin, rhodopsin could switch between the two and one would see flashes

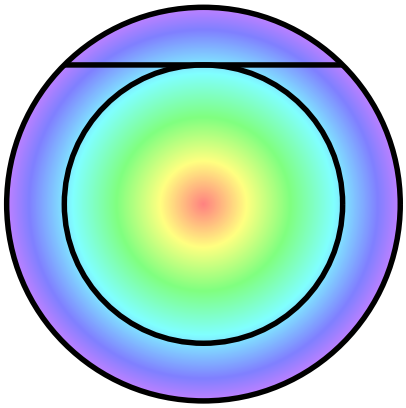

Figure 6. Velocity-potential state diagram for optimal bistate molecular machines in which $P_y = N_y$. In this condition a molecular machine cannot have an efficiency higher than ln(2) ≈ 70%. Spectral coloring suggests increasing energy with radius.

of light while in the dark, which would effectively render one blind. For EcoRI the *degenerate* sphere represents binding to and cutting any DNA sequence, which would be fatal to the bacterium. So in both cases significant intersection between the *degenerate* and *forward* spheres is eliminated by natural selection. In the high dimensional coding space, this leads to $P_y > N_y$ and ln(2) as the maximum efficiency.

Generality of coding spaces

The astute reader may have noticed that the coding spaces for rhodopsin and EcoRI appear to be constructed from different physical bases. The coding space for rhodopsin appears to be about the motion of atoms in physical space for distinguishing its coding spheres, while EcoRI has specific DNA sequences that correspond to distinct states and hence to different coding spheres. That is, there appear to be two different ways to measure the efficiency of molecular machines: successful switching to total attempts at switching ('yes/no' by rhodopsin) and information gained to energy dissipated ('info/energy' by EcoRI). While these are indeed different, the commonalities between the two systems lead to the same theoretical picture in coding space (Fig. 5 and Fig. 6), consistent with the general nature of information theory (13). First, both molecular machines, as defined previously (23), function under thermal noise and thus their parts (pins) move by approximately Gaussian distributions. Second, these parts are or can evolve to be moving independently. Since the channel capacity is reduced if there are dependencies (27), by making the parts more independent the capacity can be maximized during evolution [d_{space} increases in (**4**)].

At this point, given Gaussianicity and independence, the thermal motions of an ideal resting molecule are modeled as a sphere in a high dimensional space in both cases (23). For rhodopsin, intersecting spheres means switching states. Thermal noise could switch the state of rhodopsin so that the animal would see light when there was no photon. For EcoRI, intersecting spheres means confusion of sequences. For example instead of only binding GAATTC, thermal noise could flexibly distort the EcoRI protein so that it might also bind AAATTC, leading to inappropriate digestion of the genome ['star activity' (64)]. The coding space must map to the physical molecule, but the mapping can be different in different cases, just as an IF statement in a computer language may be supported by relays, vacuum tubes, transistors, or proteins and DNA in a genetic control circuit (65). That is, software must be supported by some physical mechanism, the hardware, but one usually cannot tell from the running software what that underlying mechanism is. Shannon's channel capacity theorem implies that both EcoRI and rhodopsin can evolve to avoid sphere intersection (confusion of states), thereby maximizing the capacity and increasing the efficiency. Any biological system having distinct states that function under thermal noise—and they all must according to the third law of thermodynamics—will have these properties. If having two distinct resting *after* states gives an advantage to the organism, then in the simplest cases the molecular machine efficiency will be maximized, evolving up to the bound of the curve shown in Fig. 2, according to equation (**7**), with the energy dissipation P_y decreasing until it just exceeds N_y. At this point the efficiency will have evolved to ln(2) ≈ 70%, as is observed.

An implication of this result is that the molecular machines must have indeed evolved to have the highest possible efficiency, as predicted by Lotka in 1922 (66). Further, since the efficiency is directly related to the channel capacity [see equation (**7**)], they must also be operating close to the maximum possible capacity. As Shannon pointed out (27) to do so they must not only have codes, but they also must be using nearly optimal codes. Recent experimental work suggests that the DNA binding protein Fis has a coding system (67) because it shows the high dimensional threshold effect predicted by Shannon (27). The observed sharp transition from binding to non-specific binding as the individual information of Fis binding sites is decreased below zero bits is apparently caused by the distinct edge of the DNA binding site recognition sphere. Similar threshold effects have been observed in restriction enzymes and other DNA recognition proteins (68). A major challenge in biology and nanotechnology is to understand what the codes of molecular machines are such that they can create sharp recognition effects and how the codes lead to the optimal efficiency of 70%.

SUMMARY

The area under a sequence logo represents the information conserved at a binding site ($R_{sequence}$). In contrast, the information needed to find the binding sites is fixed by the size of the genome and the number of binding sites required for physiological functions ($R_{frequency}$). Generally, the logo information evolves to match this required value ($R_{sequence} \rightarrow R_{frequency}$) (3, 40).

If the information in a binding site is indirectly determined by physiological functions, then how does that determine the corresponding binding energy? We can express the binding energy as the number of bits that could be gained for that energy dissipation by using a version of the second law of thermodynamics that applies when the temperature does not change, which is the case for molecular binding. This allows us to compare the actual number of bits gained by binding ($R_{sequence}$) to the maximum bits possible for the given energy dissipation (R_{energy}) to form an efficiency ($\epsilon_r = R_{sequence}/R_{energy}$) that, unlike the Carnot efficiency, applies at constant temperature. Because the energy dissipated during binding may decrease by loss of unnecessary contacts, we anticipate that most molecular systems will have evolved to a maximum efficiency. This turns out to be near 70% for a number of systems.

Following the footsteps of Bell Labs satellite engineers in 1959 (28, 29), we can use the mechanical equivalent of Shannon's channel capacity [equation (**4**)] (27) to define an equation for the isothermal efficiency [equation (**7**)] which relates it to the binding energy normalized by the thermal noise. Using the isothermal efficiency curve, we find that to explain the observed 70% efficiencies, the energy dissipated during binding must exceed the thermal noise flowing through the molecular machine at the same time.

The reason for this effect can be understood by considering a high dimension coding space. In this space the instantaneous velocity and potential energy of a molecule is represented by a point on a sphere that corresponds to the state of the molecule. The point moves by Brownian motion across the sphere and if the sphere significantly intersects another sphere, then the molecule can readily switch states. Physiology and the environment set an acceptable error rate at which inappropriate switching can occur. Inspection of the geometry of the space shows that to attain sufficient state separation only requires that the energy dissipated just exceed the thermal noise. Using the efficiency equation, this predicts a maximum efficiency of $\ln 2 = 0.69$, which is close to observed values.

When he developed information theory, Claude Shannon included a criterion which cannot be found anywhere in classical thermodynamics nor physics, namely that messages should be, and can be, chosen to be distinct (32). The equivalent concept for biological molecular machines is that molecular states can evolve to be distinct. This idea can be developed by noting that since the mechanical equivalent of voltage is the maximum potential energy (or maximum velocity) of harmonic oscillators, we can reappoint Shannon's geometrical conception of communications into the molecular situation. From that comes the important concept that it is possible to attain distinct molecular states (with a given switching error rate) if the molecules use a high enough dimension. They can do this by evolving many independent parts (pins) that vibrate as harmonic oscillators under thermal noise, which means that their velocities have Gaussian distributions (23). A combination of independent Gaussian distributions is spherical, so Shannon's message spheres correspond to distinct molecular states, also represented by spheres. Separation of states becomes easier in a high dimensional space because the surfaces of the spheres become more distinct (21, 23, 27, 60). To get from one state to another requires a velocity in a certain direction, and that corresponds to a particular rearrangement of the molecule's structure.

The concept of multiple distinct molecular states represented by spheres allowed us to steal the key prize of information theory for use in molecular biology, namely the channel capacity theorem (23). Restating the channel capacity theorem as a 'molecular machine capacity theorem', we see that because they are able to change and adapt through Darwinian evolution, biological states of molecules may become as distinct as necessary to reduce error to a level acceptable for robust survival. The molecular machine capacity theorem implies that if a system is to approach capacity it must do so by creating appropriate codes (27). So the discovery reported in this paper of 70% efficiencies leads to the additional discovery that molecular states not only can (by the molecular machine capacity theorem) but actually do evolve codes to become as distinct as necessary for survival.

ACKNOWLEDGMENTS

I thank Herbert A. Schneider (1922-2009) for continuous encouragement, and John Spouge, Martin Bier, Ilya Lyakhov, Danielle Needle, Peyman Khalichi, Carrie Paterson, Ryan Shultzaberger, Amar Klar, Peter Lemkin, Barry Zeeberg, Peter Rogan, Lynn Bayer, Zehua Chen, Blake Sweeney, Bert Gold, John Garavelli, Sorina Eftim, Mikhail Kashlev and Alex Mitrophanov for useful discussions and comments on the manuscript, and Peter Thomas and Hong Qian for pointing out the Ornstein-Uhlenbeck process to me.

FUNDING

Intramural Research Program of the National Institutes of Health; National Cancer Institute; Center for Cancer Research. Funding for open acess charge: National Cancer Institute.

Conflict of interest statement. None declared.

REFERENCES

1. Fried, M. and Crothers, D. M. (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.*, **9**, 6505–6525.
2. Crane-Robinson, C., Dragan, A. I., and Read, C. M. (2009) Defining the thermodynamics of protein/DNA complexes and their components using micro-calorimetry. *Methods Mol Biol*, **543**, 625–651.
3. Schneider, T. D., Stormo, G. D., Gold, L., and Ehrenfeucht, A. (1986) Information content of binding sites on nucleotide sequences. *J. Mol. Biol.*, **188**, 415–431 http://alum.mit.edu/www/toms/paper/schneider1986/.
4. Schneider, T. D. and Stephens, R. M. (1990) Sequence Logos: A New Way to Display Consensus Sequences. *Nucleic Acids Res.*, **18**, 6097–6100 http://alum.mit.edu/www/toms/paper/logopaper/.
5. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
6. Seeman, N. C., Rosenberg, J. M., and Rich, A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA*, **73**, 804–808.
7. Papp, P. P., Chattoraj, D. K., and Schneider, T. D. (1993) Information Analysis of Sequences that Bind the Replication Initiator RepA. *J. Mol. Biol.*, **233**, 219–230.
8. Schneider, T. D. (2001) Strong Minor Groove Base Conservation in Sequence Logos implies DNA Distortion or Base Flipping during Replication and Transcription Initiation. *Nucleic Acids Res.*, **29**, 4881–4891 http://alum.mit.edu/www/toms/paper/baseflip/.
9. Papp, P. P. and Chattoraj, D. K. (1994) Missing-base and ethylation interference footprinting of P1 plasmid replication initiator. *Nucleic Acids Res.*, **22**, 152–157.
10. Chattoraj, D. K. and Schneider, T. D. (1997) Replication control of plasmid P1 and its host chromosome: the common ground. *Prog. Nucl. Acid Res. Mol. Biol.*, **57**, 145–186.
11. Lyakhov, I. G., Hengen, P. N., Rubens, D., and Schneider, T. D. (2001) The P1 Phage Replication Protein RepA Contacts an Otherwise Inaccessible Thymine N3 Proton by DNA Distortion or Base Flipping. *Nucleic Acids Res.*, **29**, 4892–4900 http://alum.mit.edu/www/toms/paper/repan3/.
12. Shannon, C. E. (1948) A Mathematical Theory of Communication. *Bell System Tech. J.*, **27**, 379–423, 623–656 http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html.
13. Pierce, J. R. (1980) An Introduction to Information Theory: Symbols, Signals and Noise, Dover Publications, Inc., NY 2nd edition.
14. Schneider, T. D. (2010) Information Theory Primer. *Published on the web at http://alum.mit.edu/www/toms/paper/primer/,*.
15. Berg, O. G. and von Hippel, P. H. (1987) Selection of DNA Binding Sites by Regulatory Proteins, Statistical-mechanical Theory and Application to Operators and Promoters. *J. Mol. Biol.*, **193**, 723–750.
16. Stormo, G. D. (2000) DNA binding sites: representation and discovery. *Bioinformatics*, **16**, 16–23.
17. Felker, J. H. (1952) A Link Between Information and Energy. *Proc. IRE*, **40**, 728–729.
18. Schneider, T. D. (1991) Theory of Molecular Machines. II. Energy Dissipation from Molecular Machines. *J. Theor. Biol.*, **148**, 125–137 http://alum.mit.edu/www/toms/paper/edmm/.
19. Schneider, T. D. (1994) Sequence Logos, Machine/Channel Capacity, Maxwell's Demon, and Molecular Computers: a Review of the Theory of Molecular Machines. *Nanotechnology*, **5**, 1–18 http://alum.mit.edu/www/toms/paper/nano2/.
20. Carnot, S. (1824) Reflections on the Motive Power of Fire, Dover Publications, Inc., N. Y. republished in 1960, E. Mendoza, editor.
21. Callen, H. B. (1985) Thermodynamics and an Introduction to Thermostatistics, John Wiley & Sons, Ltd., N. Y. second edition.
22. Jaynes, E. T. (1988) The Evolution of Carnot's Principle. In Erickson, G. J. and Smith, C. R., (eds.), *Maximum-Entropy and Bayesian Methods in Science and Engineering*, Dordrecht, The Netherlands: Kluwer Academic Publishers http://bayes.wustl.edu/etj/articles/ccarnot.ps.gz http://bayes.wustl.edu/etj/articles/ccarnot.pdf Vol. 1, pp. 267–281.
23. Schneider, T. D. (1991) Theory of Molecular Machines. I. Channel Capacity of Molecular Machines. *J. Theor. Biol.*, **148**, 83–123 http://alum.mit.edu/www/toms/paper/ccmm/.
24. Wald, G. (1968) The Molecular Basis of Visual Excitation. *Nature*, **219**, 800–807.
25. Falk, G. and Fatt, P. (1972) Physical Changes Induced by Light in the Rod. In Dartnall, H. J. A., (ed.), *Handbook of Sensory Physiology*, Berlin: Springer-Verlag Vol. VII/1, pp. 200–244.
26. Li, P. and Champion, P. M. (1994) Investigations of the Thermal Response of Laser-Excited Biomolecules. *Biophys. J.*, **66**, 430–436.
27. Shannon, C. E. (1949) Communication in the Presence of Noise. *Proc. IRE*, **37**, 10–21.
28. Pierce, J. R. and Cutler, C. C. (1959) Interplanetary Communications. In Ordway, III, F. I., (ed.), *Advances in Space Science, Vol. 1*, N. Y.: Academic Press, Inc. pp. 55–109.
29. Raisbeck, G. (1963) Information Theory, Massachusetts Institute of Technology, Cambridge, Massachusetts.
30. Gonick, L. (1991) The Cartoon Guide to the Computer, HarperCollins, New York, NY second edition.
31. Gonick, L. and Huffman, A. (1990) The Cartoon Guide to Physics, HarperPerennial, New York, NY.
32. Schneider, T. D. (2006) Claude Shannon: Biologist. *IEEE Engineering in Medicine and Biology Magazine*, **25**(1), 30–33 http://alum.mit.edu/www/toms/papers/shannonbiologist/.
33. Heitman, J., Zinder, N. D., and Model, P. (1989) Repair of the *Escherichia coli* chromosome after *in vivo* scission by the *Eco*RI endonuclease. *Proc. Natl. Acad. Sci. USA*, **86**, 2281–2285.
34. Polisky, B., Greene, P., Garfin, D. E., McCarthy, B. J., Goodman, H. M., and Boyer, H. W. (1975) Specificity of substrate recognition by the *Eco*RI restriction endonuclease. *Proc. Natl. Acad. Sci. USA*, **72**, 3310–3314.
35. Woodhead, J. L., Bhave, N., and Malcolm, A. D. B. (1981) Cation Dependence of Restriction Endonuclease *Eco*RI Activity. *Eur. J. Biochem.*, **115**, 293–296.
36. Pingoud, A. (1985) Spermidine increases the accuracy of type II restriction endonucleases. Suppression of cleavage at degenerate, non-symmetrical sites. *Eur. J. Biochem.*, **147**, 105–109.
37. Thielking, V., Alves, J., Fliess, A., Maass, G., and Pingoud, A. (1990) Accuracy of the *Eco*RI Restriction Endonuclease: Binding and Cleavage Studies with Oligodeoxynucleotide Substrates Containing Degenerate Recognition Sequences. *Biochem.*, **29**, 4682–4691.
38. Lesser, D. R., Kurpiewski, M. R., and Jen-Jacobson, L. (1990) The energetic basis of specificity in the Eco RI endonuclease–DNA interaction. *Science*, **250**, 776–786.
39. Firsov, M. L., Donner, K., and Govardovskii, V. I. (2002) pH and rate of "dark" events in toad retinal rods: test of a hypothesis on the molecular origin of photoreceptor noise. *J. Physiol.*, **539**, 837–846.
40. Schneider, T. D. (2000) Evolution of Biological Information. *Nucleic Acids Res.*, **28**, 2794–2799 http://alum.mit.edu/www/toms/paper/ev/.
41. Kim, Y., Grable, J. C., Love, R., Greene, P. J., and Rosenberg,

J. M. (1990) Refinement of Eco RI endonuclease crystal structure: a revised protein chain tracing. *Science*, **249**, 1307–1309.
42. Lyakhov, I., Annangarachari, K., and Schneider, T. D. (2008) Discovery of Novel Tumor Suppressor p53 Response Elements Using Information Theory. *Nucleic Acids Res.*, **36**, 3828–3833 http://alum.mit.edu/www/toms/papers/p53/.
43. Benos, P. V., Bulyk, M. L., and Stormo, G. D. (2002) Additivity in protein-DNA interactions: how good an approximation is it?. *Nucleic Acids Res.*, **30**, 4442–4451.
44. Stephens, R. M. and Schneider, T. D. (1992) Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites. *J. Mol. Biol.*, **228**, 1124–1136 http://alum.mit.edu/www/toms/paper/splice/.
45. Bindewald, E., Schneider, T. D., and Shapiro, B. A. (2006) CorreLogo: An online server for 3D sequence logos of RNA and DNA alignments. *Nucleic Acids Res.*, **34**, w405–w411 http://alum.mit.edu/www/toms/papers/correlogo/.
46. Atkins, P. W. (1984) The Second Law, W. H. Freeman and Co., N. Y.
47. Clore, G. M., Gronenborn, A. M., and Davies, R. W. (1982) Theoretical Aspects of Specific and Non-specific Equilibrium Binding of Proteins to DNA as Studied by the Nitrocellulose Filter Binding Assay: Co-operative and Non-co-operative Binding to a One-dimensional Lattice. *J. Mol. Biol.*, **155**, 447–466.
48. Lehninger, A. L. (1975) Biochemistry, Worth Publishers, Inc., New York, N. Y.
49. Corning, P. A. and Kline, S. J. (1998) Thermodynamics, information and life revisited, Part II: Thermoeconomics and Control information. *Systems Research and Behavioral Science*, **15**, 453–482.
50. Schneider, T. D. (1997) Information Content of Individual Genetic Sequences. *J. Theor. Biol.*, **189**, 427–441 http://alum.mit.edu/www/toms/paper/ri/.
51. DasGupta, S., Mukhopadhyay, G., Papp, P. P., Lewis, M. S., and Chattoraj, D. K. (1993) Activation of DNA binding by the monomeric form of the P1 replication initiator RepA by heat shock proteins DnaJ and DnaK. *J. Mol. Biol.*, **232**, 23–34.
52. Schneider, T. D. (2002) Consensus Sequence Zen. *Applied Bioinformatics*, **1**, 111–119 http://alum.mit.edu/www/toms/papers/zen/.
53. Dartnall, H. J. A. (1968) The photosensitivities of visual pigments in the presence of hydroxylamine. *Vision Res.*, **8**, 339–358.
54. Schneider, G., Diller, R., and Stockburger, M. (1989) Photochemical quantum yield of bacteriorhodopsin from resonance Raman scattering as a probe for photolysis. *Chem. Phys.*, **131**, 17–29.
55. Meyer, T. E., Tollin, G., Hazzard, J. H., and Cusanovich, M. A. (1989) Photoactive yellow protein from the purple phototrophic bacterium, *Ectothiorhodospira halophila*. Quantum yield of photobleaching and effects of temperature, alcohols, glycerol, and sucrose on kinetics of photobleaching and recovery. *Biophys. J.*, **56**, 559–564.
56. Fein, A. and Szuts, E. Z. (1982) Photoreceptors: Their role in Vision, Cambridge University Press, Cambridge.
57. Tribus, M. and McIrvine, E. C. (September, 1971) Energy and Information. *Sci. Am.*, **225**(3), 179–188.
58. Keyes, R. W. (2001) Fundamental limits of silicon technology. *Proceedings of the IEEE*, **89**, 227–239.
59. Conway, J. H. and Sloane, N. J. A. (1988) Sphere Packings, Lattices and Groups, Springer-Verlag, New York.
60. Brillouin, L. (1962) Science and Information Theory, Academic Press, Inc., New York second edition.
61. Schneider, E. E., Goodeve, C. F., and Lythgoe, R. J. (1939) The spectral variation of the photosensitivity of visual purple. *Proc. Roy. Soc. Lond. A.*, **170**, 102–112.
62. Newman, A. K., Rubin, R. A., Kim, S.-H., and Modrich, P. (1981) DNA sequences of structural genes for *Eco*RI DNA restriction and modification enzymes. *J. Biol. Chem.*, **256**, 2131–2139.
63. Nathans, J. and Hogness, D. S. (1983) Isolation, Sequence Analysis, and Intron-Exon Arrangement of the Gene Encoding Bovine Rhodopsin. *Cell*, **34**, 807–814.
64. Tikchonenko, T. I., Karamov, E. V., Zavizion, B. A., and Naroditsky, B. S. (1978) *Eco*RI* activity: enzyme modification or activation of accompanying endonuclease?. *Gene*, **4**, 195–212.
65. Schneider, T. D. and Hengen, P. N. MOLECULAR COMPUTING ELEMENTS: GATES AND FLIP-FLOPS, United States Patent 6,774,222, European Patent 1057118, 2004. (2004) US Patent WO 99/42929, PCT/US99/03469. http://alum.mit.edu/www/toms/patent/molecularcomputing/.
66. Lotka, A. J. (1922) Contribution to the Energetics of Evolution. *Proc. Natl. Acad. Sci. USA*, **8**, 147–151.
67. Shultzaberger, R. K., Roberts, L. R., Lyakhov, I. G., Sidorov, I. A., Stephen, A. G., Fisher, R. J., and Schneider, T. D. (2007) Correlation between binding rate constants and individual information of *E. coli* Fis binding sites. *Nucleic Acids Res.*, **35**, 5275–5283 http://alum.mit.edu/www/toms/paper/fisbc/ http://dx.doi.org/10.1093/nar/gkm471.
68. Draper, D. E. (1993) Protein-DNA complexes: the cost of recognition. *Proc. Natl. Acad. Sci. USA*, **90**, 7429–7430.
69. Gillespie, D. T. (1996) Exact numerical simulation of the Ornstein-Uhlenbeck process and its integral. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics*, **54**, 2084–2091.

APPENDIX

In this Appendix we present an alternative geometric diagram that leads to a simple proof for why $P_y/N_y > 1$. The derivation depends on several observations about the high dimensional spheres used to model molecular states.

To recapitulate, in Shannon's 1949 model for a communications system, a series of voltage pulses are sent over a wire to form a message (27). While the transmitted pulses may be either 0 or 1 volts (for example), the received pulses vary by Gaussian distributions around each of these values because of thermal noise (23). Since the noise affects each pulse independently, the pulses can be represented by geometrically orthogonal vectors in a space with dimensions of volts. Although the transmitted message is a single point in this high dimensional space, the received message is dislocated by the thermal noise to a nearby location in the space. If a single message, given by particular pulse train, were repeated many times the received points for that message would form a sphere in the space. The assigned locations of the spheres for different messages is called the coding. Given a received set of noisy voltage pulses, represented by a single point in the space, the nearest sphere center is chosen so as to remove the noise. This process is called decoding. The challenge for designing a communications system is to place the spheres so that they do not overlap and as a consequence decoding will frequently produce the original signal. This is the key concept underlying all modern communications systems and it explains why their error rates are so low.

A similar model was developed to represent the states of molecular machines (19, 23, 32). The current state of a molecular machine is represented by the set of maximum velocities of independently moving components (pins)

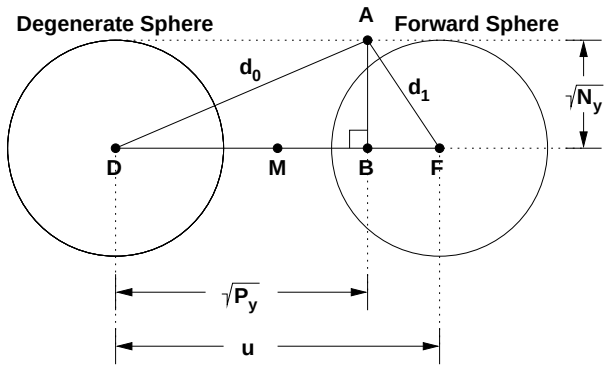

Figure 7. Molecular machine coding space configuration. Relationships between dissipated energy from a molecular machine, P_y, the thermal noise in the molecular machine N_y and the coding states are shown. Points *D* (*degenerate* sphere) and *F* (*forward* sphere) represent the attractor centers of spheres in a high dimensional coding space. *M* is the midpoint between the spheres. The distance between the sphere centers is *u* and the radius of each sphere is $\sqrt{N_y}$. Starting from the attractor centered at *D*, a molecular machine moves a distance $\sqrt{P_y}$ towards the *forward* sphere *F* to place the sphere center at *B*, but the molecule itself is deflected orthogonally by thermal noise $\sqrt{N_y}$ to the point *A*. The distance from *A* to *B* is $\sqrt{N_y}$, the distance from *A* to *D* is d_0 and the distance from *A* to *F* is d_1.

of the molecule. (The square root of the energy is proportional to the maximum velocity of an oscillator in a thermal bath.) As with the voltage model, the pins are disturbed by thermal noise and so their maximum velocities have a Gaussian distribution, which means that all of the possible movements of the molecule can be represented by a sphere in a 'velocity space', *Y* (23). Distances in this space represent changes in the shape of the molecule in a particular way, in other words, conformational rearrangements. These rearrangements can occur spontaneously if two spheres intersect.

In Shannon's model the dimensionality is presumed to be extremely large so the spherical shells are thin. Since molecules are finite, the velocity space will have a finite dimensionality so the sphere shells will have a distinct thickness (23). To avoid intersections the sphere centers may have to be separated further than twice the sphere radii. A buffer zone between the spheres reduces errors, especially in the lower dimensional spaces that biological systems may be forced to evolve in.

These considerations lay the groundwork for constructing a simple geometric diagram representing the initial ('*degenerate*') state of a molecule, placed at the origin of the velocity space and a single '*forward*' sphere placed some distance away (Fig. 7). Both spheres have radii $\sqrt{N_y}$ and the lattice spacing of the coding space is *u*, following standard conventions (59) [p. 26]. Creating a buffer zone by setting

$$u > 2\sqrt{N_y} \tag{19}$$

ensures that the fuzzy spheres have reduced intersection. The factor of 2 represents the minimum separation of the circles shown in Fig. 7, but a larger value could be used without substantially altering the proof.

The maximum velocity (potential) that the molecule has available to switch states is $\sqrt{P_y}$ (23). Suppose that the velocity $\sqrt{P_y}$ is in the direction of the *forward* sphere and sufficient to place the sphere center at point *B*, which can be inside the *forward* sphere or to the right of the midpoint *M* between the spheres. In the high dimensional space, thermal noise added to this displacement will, for the most part, be at right angles to the direction of the power P_y. (In a 100 dimensional space 99% of the noise will be at right angles to the power direction.) Thus the instantaneous state of the molecule is represented by a point *A* shown in the figure. [In Shannon's Figure 5, *A* is the 'received' point (27).]

Decoding in Shannon's voltage model consists of choosing the closest sphere center. Correspondingly, decoding in this molecular machine velocity model consists of selecting the closest sphere center by the means of an attractor around which the molecule performs noisy damped oscillation according to a multi-dimensional Ornstein-Uhlenbeck process (69). Which sphere will probably become the attractor center? Assuming that the closest sphere center will become the attractor, we can determine this by comparing two distances, d_0 the distance from *A* to *D*, the center of the *degenerate* sphere and d_1 the distance from *A* to *F*, the center of the *forward* sphere. By inspection:

$$d_0^2 = \sqrt{P_y}^2 + \sqrt{N_y}^2 \tag{20}$$
$$= P_y + N_y \tag{21}$$

and

$$d_1^2 = (u - \sqrt{P_y})^2 + \sqrt{N_y}^2. \tag{22}$$

Decoding to the *forward* sphere occurs when

$$d_1 < d_0 \tag{23}$$

from which we quickly obtain

$$\sqrt{P_y} > u/2 \tag{24}$$

by substituting equations (**21**) and (**22**) into the square of (**23**). Finally, from equations (**19**) and (**24**) we find:

$$\sqrt{P_y} > \sqrt{N_y} \tag{25}$$

or

$$P_y > N_y \quad (\mathbf{26})$$

from which a maximum efficiency of ln(2) follows directly using equation (**7**). When $P_y/N_y > 1$, decoding will usually return the molecular machine to the same state, so in the absence of power P_y the molecular machine will be stable around one attractor at a time.

All points B to the right of midpoint M decode to F. In the limit as the buffer zone is reduced at higher dimensions, $\sqrt{P_y} \rightarrow u/2 = \sqrt{N_y}$ and $B \rightarrow M$. The line segment AB can then represent the points that decode to the *forward* sphere. This corresponds to Figures 5 and 6, in which the *forward* sphere is represented by a straight line segment perpendicular to the power. As the dimension of the space increases, effectively $F \rightarrow B$, because the noise in the direction of the power is negligible, and so $d_1 \rightarrow \sqrt{N_y}$, giving the 'classical' Shannon triangle with sides $\sqrt{N_y}$, $\sqrt{P_y}$ and $\sqrt{P_y + N_y}$ (27). All points of the *forward* sphere are outside the *degenerate* sphere when $P_y > N_y$. Thus it does not matter which representation of the high dimensional space geometry is used. However, the representation of the *forward* sphere as a straight line is more appropriate when determining the radius of the volume in which point A can reside since it must be within a sphere of radius $d_0 = \sqrt{P_y + N_y}$ (23, 27). This *before* volume and the volume of the *after* spheres are used to compute the machine capacity (23).