

Self-organization and entropy reduction in a living cell

Paul C.W. Davies^a, Elisabeth Rieper^b, Jack A. Tuszynski^{c,*}

^a Beyond Center for Fundamental Concepts in Science, Arizona State University, Tempe, AZ 85287-1504, USA

^b Centre for Quantum Technologies, National University of Singapore, Block S15, 3 Science Drive 2, Singapore 117543, Singapore

^c Department of Physics, University of Alberta, Edmonton, Alberta T6G 2E1, Canada

ARTICLE INFO

Article history:

Received 22 March 2012

Received in revised form 22 October 2012

Accepted 22 October 2012

Keywords:

Genetic code

Entropy

Shannon information

Thermodynamics

Information theory

ABSTRACT

In this paper we discuss the entropy and information aspects of a living cell. Particular attention is paid to the information gain on assembling and maintaining a living state. Numerical estimates of the information and entropy reduction are given and discussed in the context of the cell's metabolic activity. We discuss a solution to an apparent paradox that there is less information content in DNA than in the proteins that are assembled based on the genetic code encrypted in DNA. When energy input required for protein synthesis is accounted for, the paradox is clearly resolved. Finally, differences between biological information and instruction are discussed.

© 2012 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Over the past century relatively little attention has been paid to the physical basis of embryology. With the discovery through the space program that the very important cytoskeletal proteins, such as microtubules, are differentially sensitive to gravity (Portet et al., 2003), it is clear that the fundamental conceptual basis of embryology cannot be developed merely by employing a classical molecular genetic framework. In addition to gravity, other physical forces such as surface tension due to intercellular adhesion (Goel et al., 1970, 1975; Gordon et al., 1972, 1975; Steinberg, 1996; Trinkaus, 1984a), the mechanical forces exerted during cell division (Rappaport, 1996), and physical waves of cytoskeletal expansion and contraction that traverse embryos (Gordon, 1999) all provide important non-chemical contributions to morphogenesis of embryos. Moreover, cell differentiation leading to morphological differences between cells in various tissues and organs are extreme examples of what appears to be an entropy reduction process, i.e. self-organization. This apparent contradiction of the second law of thermodynamics drew attention of many physicists beginning with Schrödinger (1967). At the macroscopic level, living systems are thermodynamically open and far-from-equilibrium systems, hence the balance of entropy at this level must necessarily involve metabolic energy production as well as heat and waste product dissipation into the external environment. A more subtle

question arises at the level of a single cell, especially an embryonic cell, that undergoes rapid self-organization, all the while being sensitive to physical forces acting on it from the environment. Houlston et al. (1993) have shown by means of time-lapse video recordings that a wave of cytoplasmic reorganization, involving displacement of perinuclear organelles and movements of the surface relative to the underlying layer of cytoplasm, occurs prior to first cleavage. This wave requires dynamic microtubules in association with the organelles. Precleavage waves have also been described, progressing in the same direction (Yoneda et al., 1982). In many species rearrangements of the cytoplasm directed by external cues may introduce new axes of cleavage (Driesch and Morgan, 1895; Speksnijder et al., 1990). Thus, there appears to be a relationship between microtubules and rearrangements of cytoplasmic components induced by precleavage waves, but the correlation between them is unknown. It is, therefore, important to understand the entropy reduction contributions arising from internal self-organization, information storage and transfer, at a single cell level, as the only way to reconcile this with laws of thermodynamics is by balancing these free energy changes with metabolic energy expenditures. We will discuss these processes in some detail in this paper. This paper is largely aimed at providing an overview of the problem of linking biological organization with information and entropy.

Living cells perform numerous complicated, synchronized and very specific tasks in order to maintain their biological functions. These complex tasks require information input (e.g. a chemical gradient), information processing (signals sent into the cell from membrane receptors), and instruction (e.g. reorganization of the actin cytoskeleton for motility) as the output of the underlying

* Corresponding author at: 3-336 Cross Cancer Institute, 11560, University Avenue, Edmonton, Alberta T6G 1Z2, Canada. Tel.: +1 780 432 8906.

E-mail addresses: jackt@ualberta.ca, jack.tuszynski@gmail.com (J.A. Tuszynski).

computations. In order to function, a living cell, similarly to a man-made machine, requires specific components to be interconnected in an intelligent fashion so they can perform the desired tasks. In addition, a steady supply of energy must be provided to be converted, with some level of efficiency, into useful work and to keep the organism at a constant physiological temperature. However, it should be noted that living organisms cannot simply be reduced to machines as demonstrated by Rosen (1991). In fact, organisms are different from machines because they are characterized as being closed to efficient cause which means that the catalysts needed for its operation must be generated internally. As Rosen stated it “A material system is an organism if, and only if, it is closed to efficient causation”. The closure of the relational diagram he showed establishes a category of objects called organisms that are clearly distinguishable from machines. This distinction arose from a procedure which did not reduce the system to its material parts, nor did it explicitly invoke dynamics. Also, the concept of replication in this context means that what is replicated is a functional component, not a material part as such. Thus while organisms are complex, not all complex objects are organisms. An organism possesses the kind of *unity* invoked when discussing autopoietic systems (Maturana and Varela, 1980). It is necessary and useful to recognize functional components as making up separate tangible aspects of the system. Biological cells are constructed from yet smaller machine-like entities called organelles. Cell organelles include mitochondria, Golgi complexes, endoplasmic reticulum, and the protein filaments of the cytoskeleton such as microtubules and actin filaments (microfilaments). Even below this level there are machine-like parts of the cell, namely motor proteins and enzymes, that perform specific functions involving energy input and power output, e.g. transport, motility and cell division (Alberts et al., 1994). A critically important macromolecule is ATP (and its relative GTP), which serves as the primary energy currency of the cell. ATP is used to build complex molecules, provide energy for nearly all living processes in order to power virtually every activity of the cell. While organic components of nutrients contain numerous low-energy covalent bonds, they are not directly useful to do most type of work in the cell. Thus, low energy bonds must be translated into high-energy bonds using ATP energy by removing one of the phosphate-oxygen groups, turning ATP into ADP. Subsequently, ADP is usually immediately recycled in the mitochondria where it is recharged and re-emerges again as ATP. ATP synthesis in a mitochondrion requires approximately 60 kJ/mol of energy delivered through complex and well-tuned electron transport reactions. ATP hydrolysis releases approximately 30.5 kJ/mol of free energy (dependent on the concentration and pH values), which can be viewed as a biological energy unit. The human body requires the production of its weight in ATP every day in order to function, which translates into 10^{21} ATP molecules per second. Since there are on the order of 3.5×10^{13} cells in the human body and each cell has on the order of 10^3 mitochondria, there are approximately 3×10^4 ATP production events per mitochondrion per second. This process involves a complex set of biochemical reactions called oxidative phosphorylation whose net effect is a conversion of one molecule of glucose into 38 molecules ATP. At any instant each cell contains about one billion ATP molecules. Because the amount of energy released in ATP hydrolysis is very close to that needed by most biological reactions, little energy is wasted in the process. Generally, ATP is coupled to another reaction such that the two reactions occur nearby utilizing the same enzyme complex. Release of phosphate from ATP is exothermic while the coupled reaction is endothermic. The terminal phosphate group is then transferred by hydrolysis to another compound, via a process called phosphorylation, producing ADP, phosphate (Pi) and energy. Phosphorylation often takes place in cascades becoming an important signaling mechanism within the cell. Importantly, ATP is not excessively unstable, but it is designed

so that its hydrolysis is slow in the absence of a catalyst. This insures that its stored energy is released only in the presence of an appropriate enzyme. The mitochondrion, where ATP is produced, itself functions to produce an electro-chemical gradient – similar to a battery – by accumulating hydrogen ions between the inner and outer membrane. This electro-chemical energy comes from the estimated 10,000 enzyme chains in the membranous sacks on the mitochondrial walls. As the charge builds up, it provides an electrical potential that releases its energy by causing a flow of hydrogen ions across the inner membrane into the inner chamber. The energy causes an enzyme to be attached to ADP, which catalyzes the addition of a third phosphorus to form ATP. Energy production and utilization is essential to life and is also part of the information processing equation as will be discussed later in the paper.

2. Probability, entropy and information

From the point of view of information processing, the description of both a complex system with a large number of degrees of freedom and a system with a small number of degrees of freedom that is unstable, practically requires an infinite number of bits of information. Biological systems fall into this category and require a statistical description of their behavior. Statistical physics (Penrose, 1979) offers a simple method to circumvent the problem with incomplete knowledge of the system's initial state. Instead of a single system, it concerns itself with an ensemble of many identical copies of the same system (called replicas) that only differ in the choice of their initial state. The state of an ensemble with a uniform distribution of states over the available domain in phase space of the individual system is referred to as thermodynamic equilibrium. Given the volume of the available domain \mathbf{S} in the phase space we assign the meaning of probability $P(\mathbf{A})$ to the volume of an arbitrary subset \mathbf{A} of \mathbf{S} . If domain \mathbf{S} is not covered uniformly by its states, then we introduce the probability density $\rho(s)$ determined for all states s in \mathbf{S} . The probability that a state belongs to a given subset \mathbf{A} of the available state space \mathbf{S} , is

$$P(\mathbf{A}) = \int_{\mathbf{A}} ds \rho(s). \quad (1)$$

The probability density is normalized to unity, so that

$$\int_{\mathbf{S}} ds \rho(s) = 1. \quad (2)$$

If the probability $P(\mathbf{A})$ is known, then as a result of finding that the state indeed belongs to the set \mathbf{A} the observer has gained a certain amount of information about the system under consideration and what was initially a probability $P(\mathbf{A})$ has now become a certainty. The smaller the probability $P(\mathbf{A})$, the greater the information gain. Conversely, if $P(\mathbf{A})$ was large, the information gain is small. Hence, the information gain I is a decreasing function of P . The amount of information about two independent events whose probability is the product of individual probabilities is the sum of the information values for each of the events separately. A function of probability P that has both these properties can be defined as

$$I(A) = -k \log_b P(\mathbf{A}), \quad (3)$$

where k and b are constants. Since the logarithm of unity is zero, the outcome of an event \mathbf{A} that was certain before the observation took place, $P(\mathbf{A}) = 1$, and no new information is gained, $I = 0$. Eq. (3) was first derived in a famous paper on information theory (Shannon, 1948). The coefficients k and b determine the units of information. For $k = 1$ and $b = 2$ we obtain the unit of information called one bit. One bit is the amount of information gained as a result of an

experiment with two equally probable outcomes ($P=1/2$), which can be represented by the binary digits 0 or 1:

$$I = -\log_2 \frac{1}{2} = 1. \quad (4)$$

If, in addition to the probability $P(\mathbf{A})$ that a state belongs to a certain subset \mathbf{A} of the state space \mathbf{S} , the entire probability density $\rho(s)$ is known, we can determine the average information gain achieved by the observer after the determination of state s of the system as

$$S = -k_B \int ds \rho(s) \ln \rho(s). \quad (5)$$

This expression is identical to the one first proposed by Boltzmann (1872) and later extended to a general situation of an arbitrary statistical ensemble by Gibbs (1902). These expressions provided a statistical interpretation of entropy, a quantity whose term was coined by Clausius (1865). Entropy represents the part of the internal energy that cannot be used for work, divided by absolute temperature T . Here, \ln stands for the natural logarithm with a base $b=e \approx 2.718$, while the constant k was selected as $k_B = 1.38 \times 10^{-23}$ J/K, which equals to the gas constant R divided by Avogadro's number N_A and is called the Boltzmann constant. The unit of entropy is therefore J/K.

For a given domain of available states in the phase space, the probability density which is constant in this domain \mathbf{A} , and zero elsewhere (corresponding to the microcanonical ensemble), is given by

$$\rho(s) = \begin{cases} \Omega^{-1} & \text{for } s \in \mathbf{A} \\ 0 & \text{for } s \notin \mathbf{A} \end{cases}. \quad (6)$$

For the probability to be properly normalized to unity, the constant Ω represents the volume of domain \mathbf{A} : $\Omega = \int_{\mathbf{A}} ds$. The entropy is then

$$S = -k_B \int_{\mathbf{S}} ds \rho(s) \ln \rho(s) = k_B \int_{\mathbf{A}} ds \Omega^{-1} \ln \Omega = k_B \ln \Omega. \quad (7)$$

The instability of motion, i.e. the process of mixing on a microscopic scale, leads to the law of entropy increase over time, which is commonly referred to as the second law of thermodynamics.

Mawell (1871) contemplated the concept of a hypothetical being that would be able to observe the velocities of individual gas molecules moving about in a container. The container would have a partition with an opening that can be covered by a latch. This "thought experiment" has been referred to as Maxwell's demon since it was imagined that a hypothetical demon would be in charge of closing and opening the hole in the partition allowing only sufficiently fast particles to move from right to left and only sufficiently slow ones to move from left to right across the partition. This would over time result in a temperature increase in the left part of the container and a temperature decrease in the right part of the container. The thus created temperature gradient is in contradiction with the second law of thermodynamics due to the work performed in the process by thermal fluctuations alone in a gas at a thermodynamic equilibrium.

Shannon's information defined as negative entropy has enabled the resolution of the Maxwell's demon paradox in terms of the energy cost put on information content. Szilard's solution (Szilard, 1929) of the problem endowed the demon with information, i.e. Shannon's information balancing out the changes in the entropy of the gas. The energetic cost of one bit of information at physiological temperature is $\ln 2 k_B T = 3 \times 10^{-21}$ J = 18.5 meV which is comparable to the thermal energy unit $k_B T$ (Cook, 1984) freely available in a system kept at a constant temperature T . While a single bit of information has almost no cost in terms of energy input, several bits require concerted effort in order to be encoded.

3. Nucleic acids, DNA, and the genetic code

Nucleic acids are used to store and transmit genetic information in cells and come in two varieties: the deoxyribonucleic acid (DNA) and the ribonucleic acid (RNA). RNA differs from DNA by having one additional oxygen atom on each sugar and one missing carbon atom on each thymine base. DNA and RNA perform distinct functions in the cell. DNA is more stable and hence better suited for information storage. RNA is less stable and hence more useful in information transfer as a messenger, a translator and a synthetic machine (Saenger, 1984).

Nucleic acids are linear polymers of nucleotides, each of which is composed of a sugar phosphate group and a disk-shaped base group linked by phosphodiester bonds $-\text{O}-\text{PO}_2^--\text{O}-$. The sugar ribose occurs only in RNA. In DNA, it is replaced by the sugar deoxyribose. The sugar-phosphate groups are connected together to form a hydrophilic backbone (phosphates have a negative electric charge) and the mainly hydrophobic bases are located off the side of the chain. In a water environment, one side of the chain is protected by the hydrophilic backbone while the other side is exposed. The formation of a phosphodiester bond is an endoergic reaction and needs free energy, which is released in the nucleotide triphosphate hydrolysis. The sequence in which the individual nucleotides of side chains occur along the nucleic acid chain is strictly fixed and genetically determined. In DNA there are 4 "canonical" bases, guanine (G) and adenine (A), being derivatives of pyrimidine, and cytosine (C) and thymine (T), being derivatives of purine. In RNA thymine is replaced by uracil (U). The edges of these bases are chemically complementary such that A forms two hydrogen bonds with T, and C three hydrogen bonds with G. No other combinations of pairs lead to bond formation between them. The matching patterns of hydrogen bonds allow a second strand of DNA, provided it has the proper base sequence, to form a stable complex, the famous double helix. In a double-stranded DNA molecule, its bases lie in the interior of the helix and are held together by hydrogen bonds.

All living organisms pass on genetic information from generation to generation. This information is contained in the chromosomes, which are made up of genes. The information needed to produce a particular type of protein molecule is contained within each gene. Genetic information contained within a gene is built into DNA. The arrangement of A paired to T and G opposite C ensures that genetic information is passed to the next generation accurately. The two strands or helices of the DNA separate with the help of enzymes, leaving the charged parts of the bases fully exposed (Sinden, 1994).

Genetic coding can be discussed using information theory. The so-called genetic code is a detailed prescription regarding the synthesis of proteins according to the algorithm given in Fig. 1.

The four nucleic acids (U, C, A, G or T, C, A, G, respectively) can be ordered in triples, termed codons, in RNA or DNA to form a genetic code for transcription and translation into amino acids. While there are 64 possibilities of forming triple sets of the RNA or DNA bases, there are numerous redundancies such that there is correspondence to the 20 naturally occurring standard amino acids, a stop codon (three combinations of amino acids code for a stop), and a start codon (which also codes for the amino acid methionine), giving rise to 22 elementary building blocks (distinct pieces of information) for the construction of proteins.

Specific sequences of nucleotides code for the specific sequences of amino acids that form proteins. Thus, discovering that any one of these is in a specific location along a strand increases the information content by $\log_2 4 = 2$ bits. However, as there are 20 amino acids, the identification of one particular amino acid requires $\log_2 20 \approx 4.22$ bits of information. We therefore conclude that coding must be done by at least three nucleotides arranged in order. Later on we discuss if information can be gained in the process of

		Second Position							
		U		C		A		G	
First Position		code	Amino Acid	code	Amino Acid	code	Amino Acid	code	Amino Acid
		U	UUU	phe	UCU	ser	UAU	tyr	UGU
UUC			UCC	UAC			UGC		C
UUA	leu		UCA	UAA	STOP		UGA	STOP	A
UUG			UCG	UAG	STOP		UGG	trp	G
C	CUU	leu	CCU	pro	CAU	his	CGU	arg	U
	CUC		CCC		CAC	CGC	C		
	CUA		CCA		CAA	gln	CGA		A
	CUG		CCG		CAG	CGG	G		
A	AUU	ile	ACU	thr	AAU	asn	AGU	ser	U
	AUC		ACC		AAC	AGC	C		
	AUA		ACA		AAA	lys	AGA	A	
	AUG		ACG		AAG	arg	AGG	G	
G	GUU	val	GCU	ala	GAU	asp	GGU	gly	U
	GUC		GCC		GAC	GGC	C		
	GUA		GCA		GAA	glu	GGA		A
	GUG		GCG		GAG	GGG	G		

Fig. 1. The genetic code.

translation, which transfers information from an RNA sequence to an amino acid sequence of a protein.

In a biological cell the DNA is too valuable to do the work of synthesizing proteins directly. Transcription of the coded DNA is accomplished by producing from it what are called messenger RNA (mRNA) molecules. The production of proteins is directed by the mRNA, when they travel to the ribosomes, by specifying the sequence in which amino acids will be linked. We have seen that three of the 64 possible combinations do not code for any amino acid and act as terminators. When the ribosome reaches this portion of the mRNA chain, the growth of the protein is halted. If a mutation or error occurs in the DNA, all copies will carry it and incorrect synthesis will take place at the ribosomes, sometimes leading to diseases and pathological abnormalities.

The sequence of events: DNA → mRNA → polypeptide → folded protein is illustrated schematically in Fig. 2.

There are 20 “canonical” amino acids; some others occur, but very rarely (Stryer, 1981). Using the Swiss-Prot and GenBank databases as the basis for the calculations, a probability table has been constructed (Shen et al., 2006) for the occurrence of the 20 naturally occurring amino acids (see Table 1).

In the 20 amino acids designed by nature, what varies is the side chain. Some side chains are hydrophilic while others are hydrophobic. Since these side chains stick out from the backbone of the molecule, they help determine the properties of the protein made from them. The side chains exhibit a wide chemical variety and can be grouped into three categories: non-polar, uncharged polar, and charged polar. The sequence of amino acids in each polypeptide or protein is unique giving its characteristic three-dimensional shape or native conformation.

4. Amino acids, peptides, and proteins

In order to form a polymeric chain, amino acids are condensed with one another through dehydration synthesis reactions, which are not spontaneous but occur through the energy-driven action

of the ribosome. A polypeptide chain of amino acids consists of a regularly repeating part called the main chain or the backbone, and a variable part, consisting of distinctive side chains. In some cases even a change in one amino acid in the sequence can alter the protein's ability to function. The formation of a peptide bond is an endoergic reaction and needs free energy, which is usually released in the guanosine triphosphate (GTP) hydrolysis.

Long amino acid chains form proteins, which are the most abundant macromolecules in the living cells, present in their membranes, cytosol, cell organelles, and chromosomes. They constitute about 50% of the dry mass of a living cell. Each type of protein has its own unique structure and function. Proteins, due to their diversity and functionality, perform most of the typical tasks of the cell.

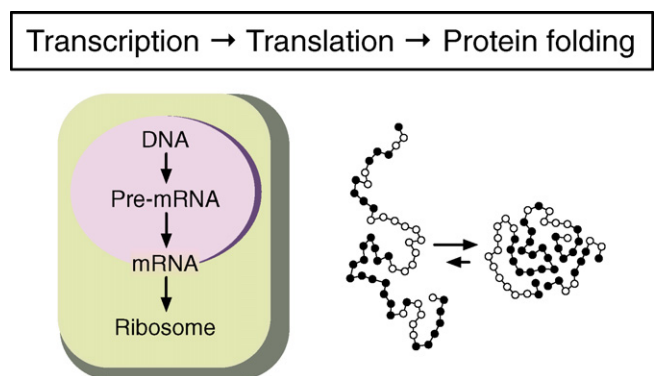


Fig. 2. Processing of genetic information. Classical dogma: the information is carried by DNA, which undergoes replication during the process of reproduction and transcription into RNA when it is to be expressed; gene expression consists in translation of information written on RNA onto a particular protein primary structure. The polypeptide consisting of specific amino acids is folded into the protein's tertiary structure. The left panel represents a cell, and depicts the information flow from DNA, through mRNA, to translation at a ribosome. The right panel represents the transition between the unfolded and folded states of a protein.

Table 1

Frequency distribution of all 20 naturally occurring amino acids in annotated protein sequences contained in the Swiss-Prot database (release 43.0), and the frequency distribution of the amino acids in characterized and hypothetical human protein sequences from in the National Center for Biotechnology Information (NCBI) GenBank database (Build Number 34). The occurrence of each individual amino acid has been averaged over the total number of amino acids found within each of the databases. Table reproduced from Shen et al. (2006).

Amino acid single letter code	Swiss-Prot frequency	GenBank frequency
A	0.0777	0.0704
C	0.0157	0.0231
D	0.0530	0.0484
E	0.0656	0.0692
F	0.0405	0.0378
G	0.0691	0.0675
H	0.0227	0.0256
I	0.0591	0.0450
K	0.0595	0.0565
L	0.0960	0.0984
M	0.0238	0.0237
N	0.0427	0.0368
P	0.0469	0.0610
Q	0.0393	0.0465
R	0.0526	0.0552
S	0.0694	0.0799
T	0.0550	0.0534
V	0.0667	0.0613
W	0.0118	0.0121
Y	0.0311	0.0282

All structural and functional properties of proteins derive from the chemical properties of the polypeptide chain.

The levels of structural organization in globular proteins can be listed as follows:

Primary sequence → secondary sequence → supersecondary sequence → domain → globular protein → aggregate

In a nutshell, the primary structure is defined as the linear sequence of amino acids in a polypeptide chain. It represents the covalent backbone and the linear sequence of the amino acid residues in the peptide chain. The secondary structure is characterized by spatial organization in terms of structural motifs such as alpha helices, beta sheets, random coils, triple helices and an assortment of other motifs usually in localized regions of a protein. Further folding of polypeptide chains results in the formation of tertiary structures that result from long-range contacts within the chain that are stabilized by a combination of van der Waals, hydrogen, electrostatic, and disulfide bonds. The quaternary structure is the organization of protein subunits, or two or more independent polypeptide chains, and it represents aggregates in which globular proteins are bound by non-covalent interactions spontaneously forming oligomers of varying sizes.

The spontaneous act of protein folding is remarkable in that the complex motion of the protein's structural elements, i.e. amino acids, transfers the one-dimensional sequence of data into a three-dimensional object and the process is a result of thermally activated Brownian motion. The net stabilization of the native state conformation of a protein results from the balance of large forces that favor both folding and unfolding. The net free energy of folding is 42 kJ/mol.

To estimate how long it would take a 100-residue polypeptide to complete a random search for the native state we assume there are three possible conformational states for each residue and that it takes 10^{-13} s to interconvert between each state. For the 100-residue polypeptide, there are 3^{100} [$=5 \times 10^{47}$] possible conformational states. Assuming a single unique native state conformation, it would take 5×10^{34} s or 1.6×10^{27} y. This absurd result, often referred as the Levinthal Paradox, clearly shows that protein folding does not occur by random search.

A protein can assume a very large number of related but distinct conformational states whose distribution can be described by a so-called energy landscape. Each substate is a valley in a $3N - 6$ dimensional hyperspace where N is the number of atoms forming the protein. The energy barriers between different substates range from about 0.2 kJ/mol to about 70 kJ/mol. Note that conformational entropy is meant to define the entropy associated with the multiplicity of conformational states of the disordered polypeptide chain.

Strait and Dewey (1996) analyzed a large data set of proteins to determine the Shannon information content of a protein sequence. In a statistical database, the most important issue is to properly define the phase space and specifically whether the elements follow Markovian or non-Markovian interaction type. This information entropy of proteins was estimated by three distinct methods, namely a k -tuple analysis, a generalized Zipf analysis, and a "Chou-Fasman gambler." Each method results in somewhat different information values. The k -tuple analysis is a "letter" analysis, based on conditional sequence probabilities. The generalized Zipf analysis uses statistical linguistic qualities of protein sequences to determine the Shannon entropy. The Zipf analysis and k -tuple analysis estimate the values of Shannon entropies as 2.5 bits/amino acid. This is much smaller than the value of 4.18 bits/amino acid obtained from the nonuniform composition of amino acids in proteins. The "Chou-Fasman" gambler algorithm is based on the use of specific rules for protein structure generation. It uses both sequence and secondary structure information to predict the number of possible amino acids that could form a proper sequence. The information content of the 3D structure of a protein was calculated using the Kolmogorov entropy (Dewey, 1996) which is defined as the length in bits of the shortest computer program required to describe a given object. Dewey (1996) showed that the Kolmogorov entropy of a protein was less than 1.0 bits/amino acid. Thus, there is more than enough information available in the protein sequence to determine the protein structure. While the number of possible protein sequences is greater than immense number $I = 10^{110}$, there are correlations, which lead to grouping and redundancy. This phase space reduction aspect, or compartmentalization, controls the structure building in the cell. Pande et al. (1994) discussed the role of non-randomness in protein sequences with implications for evolution. Compartmentalization of sequences including applications of multi-fractals was further analyzed by Dewey and Strait (1996).

5. Entropy reduction in living systems

Living cells are dissipative, open, and far-from-equilibrium systems that lower the entropy utilizing an influx of energy and molecular material in a multi-compartment structure with specific functional characteristics. Entropy reduction was discussed early on by Schrödinger (1967) and it relies on both energy supply to create a metastable non-equilibrium state and electrical, pressure and chemical potential gradients across semi-permeable membranes. Electric potential differences also assist in the process. As an open system, a cell operates cyclically exchanging material and heat with the environment. High-energy molecules are absorbed through pores in the membrane and their energy used to synthesize components of the cell and maintain ambient temperature. Heat is dissipated and waste products excreted so that excess entropy in the environment is balanced by structure- and information-production lowering the entropy inside the cell. This, of course, leads to a net entropy change in the cell fluctuating quasi-periodically close to the zero value. Cell death would manifest itself in the breakdown of structures and functions leading to a continuous entropy production as governed by the second law

of thermodynamics. Overall, the entropy changes in the cell can be attributed to four distinct processes: (a) chemical reactions leading to the aggregation of molecules, (b) mass transport in and out of the cell leading to concentration gradients across the membrane, (c) heat generation due to metabolism of the cell, and (d) information stored in terms of genetic code in both nuclear and mitochondrial DNA. Morowitz (1955) estimated that approximately 2×10^{11} bits of information are contained in the structure of *Escherichia coli* bacteria, the simplest and best documented organism, a number which agrees with calorimetric data (Gilbert, 1966). However, the estimated information capacity in the *E. coli*'s genome is only 10^7 (Johnson, 1970), which is at first surprising but on closer examination, to be expected, as argued below.

Living cells, as all matter, must obey the energy conservation principle, which takes the form of the first law of thermodynamics

$$dU = DQ + DW. \quad (8)$$

In the thermodynamic sense, cells can be viewed as machines, similar to a combustion engine engaged in a Carnot cycle, performing work and generating heat, requiring constant supply of energy and matter (i.e. energy-giving molecules like glucose). A more appropriate formulation of the energy balance is through the Gibbs free energy that accounts for a change in the numbers of molecules and the presence of several molecular species:

$$G = U - TS + PV = \mu N \quad \text{or} \quad dG = -SdT + VdP + \mu dN. \quad (9)$$

Hence, the entropy differential can be written as

$$dS = \frac{dU}{T} + \frac{PdV}{T} - \frac{\mu dN}{T}, \quad (10)$$

which indicates that entropy changes can be achieved through heat production, change of volume or a flux of molecules.

Since the entropy of an ideal gas of N particles with total energy E , of mass m each, is (Landau and Lifshitz, 1969)

$$\frac{S}{k} = N \left\{ \ln \left(\frac{V}{N} \right) + \frac{3}{2} \ln \left(\frac{mE}{3\pi\hbar^2 N} \right) + \frac{5}{2} \right\} \quad (11)$$

this means that confining molecules within space, as is the case with building a cellular structure, reduces the exploration volume V , and thus reduces the entropy of the system accordingly. Conversely, mixing two molecular species with numbers N_1 and N_2 in a fixed volume V by opening a partition between their compartments V_1 and V_2 increases the entropy by the amount given below:

$$\Delta S = N_1 \ln \left(\frac{N}{N_1} \right) + N_2 \ln \left(\frac{N}{N_2} \right). \quad (12)$$

Therefore, keeping various molecular species separated in individual compartments (including the mitochondria, the nucleus, the endoplasmic reticulum, etc.) is another entropy reducing process. While the above equations strictly speaking apply to equilibrium situations, an assembled structure of the cell from its building blocks, by and large stays in its morphological state except for mitosis and continuous material transport which can be regarded as a second order correction.

Marina et al. (2009) derived specific equations describing the entropy due to the compartmentalization of components in eukaryotic cells as a function of cell and compartment volumes, and of the concentration of solutes. Based on both known and estimated values of volume and solute concentrations they found that the contribution of compartmentalization to the decrease of entropy is approximately -14.4×10^{-14} J/K cell (-0.7 J/KL) in the case of *Saccharomyces cerevisiae*, a typical eukaryotic cell, and approximately -49.6×10^{-14} J/K cell (-1.0 J/KL) in the more complex *Chlamydomonas reinhardtii*. They compared these values with other possible contributions to entropy reduction, such as the informational entropy of DNA and the conformational entropy of proteins.

They concluded that compartmentalization is the most essential development that significantly decreases the entropy of living cells during biological evolution.

Enzymatic catalysis against the energy barrier is a process that helps achieve such a deliberate separation of molecular species. In fact, a variety of solute molecules are contained within cells. The cellular fluid (cytosol) has a chemical composition of 140 mM K^+ , 12 mM Na^+ , 4 mM Cl^- and 148 mM A^- , where the symbol A stands for protein. Cell walls are semipermeable membranes and permit transport of water but not of solute molecules. We use Dalton's Law to determine the osmotic pressure inside a cell. A mixture of chemicals, with concentrations c_1, c_2, c_3, \dots , dissolved in water has the total osmotic pressure equal to the sum of the partial osmotic pressures, Π , of each chemical: $\Pi = \Pi_1 + \Pi_2 + \Pi_3 + \dots = RT(c_1 + c_2 + c_3 + \dots)$. The total osmotic pressure inside a cell, Π_{in} , can be estimated as 7.8×10^4 Pa while the cell exterior is composed of 4 mM K^+ , 150 mM Na^+ , 120 mM Cl^- and 34 mM A^- and, as a consequence the total osmotic pressure of the cell exterior, Π_{out} , is given by 7.9×10^4 Pa. Because Π_{in} and Π_{out} are quite close in value, the osmotic pressure difference between the exterior and interior part of the cell is very small, as it is the net pressure exerted on the cell wall that matters. The cell has a sophisticated control mechanism to do this. This can, again, be seen as an entropy reduction mechanism.

Looking deeper into the issue of entropy reduction by the cellular process, in the production of macromolecules such as proteins, naturally the atoms that are assembled lose their degrees of freedom by being joined together. In the simplest case of a peptide chain viewed as a semi-flexible rod, each amino acid prior to the assembly process possesses three translational and three rotational degrees of freedom, in addition to some internal degrees of freedom, which by and large survive the assembly process. After a peptide has been assembled, only small rotations around the backbone are permitted effectively eliminating five degrees of freedom per amino acid. Consequently, this can be viewed as an entropy reduction process. This negative conformational entropy is created in addition to the combinatorial contribution that described the probability of selecting a particular sequence of amino acids in the peptide, i.e. $k \ln 20^N$ where N is the number of amino acids in a peptide. The folding of a chain into a globular protein restricts the motion of its member groups eliminating some rotations altogether and limiting others. This, again, can be seen as a reduction of the phase space whose volume changes from Ω to Ω' with an attendant entropy reduction of $\Delta S = k \ln(\Omega/\Omega')$. For illustration purposes, we have used here a somewhat simplistic approach via a microcanonical ensemble where all states in the phase space have the same probability while in reality (see Table 1), a canonical ensemble should be used leading to a more accurate but also a more complicated formula, namely:

$$S = k_B T \frac{\partial}{\partial T} \ln Z + k_B \ln Z, \quad (13)$$

where $Z = \sum_i \exp(-E_i/kT)$ is the partition function of the system. However, since the system is open, a grand canonical ensemble technique should in fact be used in the evaluation of the resultant entropy change.

The protein synthesis process, when looked at from the point of view that starts with the information content in the coding regions of DNA and continues with transcription into RNA followed by translation into an amino acid sequence, brings another entropy reduction paradox into light. Suppose a particular gene has $3N$ nucleotides and codes for a protein with N amino acids. Because there are four types of nucleotides available and 20 amino acids to choose from, the corresponding entropies do not match. The combinatorial entropy of this particular gene can be calculated as $S_1 = 3Nk_B \ln 4$ while that of the resultant protein is $S_2 = Nk_B \ln 20$.

The difference between the final (protein) entropy and the initial (gene) entropy is always negative since:

$$\Delta S = S_2 - S_1 = Nk_B \ln \left(\frac{20}{64} \right) \quad (14)$$

or in terms of the equivalent molar heat of reaction:

$$Q = T\Delta S = -1.16RT. \quad (15)$$

Clearly, a spontaneous process that cools the environment contradicts the second law of thermodynamics. However, a process that requires the input of work to reduce entropy (and hence expelling heat to the environment while cooling the local area where the reaction takes place) is well-known. Refrigerators operate on exactly the same principle. The question is: “Where does energy input come into play in protein synthesis?” It turns out that the energy cost of protein synthesis is very substantial and can be summarized as following these steps:

1. Charging of tRNA requires the input of two ATP molecules.
2. Binding of tRNA to a ribosome requires the input of one GTP molecule.
3. Translocation requires the input of one GTP molecule.

The total cost can be estimated as the energy of four high-energy phosphate bonds for each peptide bond formed, i.e. per each amino acid polymerized. This does not include additional energy costs involved in DNA transcription. Since the free energy released in the hydrolysis of ATP into ADP amounts to approximately 30.5 kJ/mol, and the hydrolysis of GTP is highly substrate dependent but comparable, we conclude that the work that needs to be performed by the cell in the process of adding an amino acid to a peptide sequence in the faithful performance of protein synthesis is at least 120 kJ/mol, or

$$\Delta W = 48RT, \quad (16)$$

which exceeds by almost two orders of magnitude the corresponding entropy reduction contribution discussed above. Other aspects that can be included in the total entropy analysis involve the change in the translational entropy of water that surrounds both DNA and protein surfaces and, as a result, loses several degrees of freedom per molecule. It has been demonstrated that a folding process such as protein folding leads to a large amount of entropy increase (Kinoshita, 2009) which would additionally mediate the effects of combinatorial entropy reduction. In conclusion, when precise energy balance is made, the seeming paradox of entropy reduction is clearly resolved by the energy cost of the cellular machinery that enzymatically catalyzes the required biochemical reactions taking place (Lambert, 1984).

It should be stressed that the second law of thermodynamics is valid for closed systems, and it states that in closed systems irreversible processes such as heat generation lead to entropy increases while reversible processes involve no heat and no entropy change. Through the introduction of mean values of chemical potentials, the second law of thermodynamics has been generalized to open systems. In chemical thermodynamics, it is common to compute equilibrium composition at constant values of the chemical potential and it is done the same way by minimizing the corresponding potential function. A living cell is an open system, and taken together with its surroundings the total entropy change should never be negative. In closed systems conditions for equilibria are expressed as either minima of the appropriate thermodynamic potentials (e.g. Gibbs free energy) or maximum entropy requirements (Landau and Lifshitz, 1969). In open systems there is no such rule, except one looks for stability conditions of a given state, i.e. whether under a small perturbation the state will evolve or retain its equilibrium value.

Another way of discussing entropy is in terms of order and disorder. The most pertinent physical transformations between states of matter that are ordered and disordered are called phase transitions. Continuous (second order) phase transitions involve no entropy change at the critical point, and ordering in the system sets in gradually, as seen through the bifurcation of an associated order parameter. In first order phase transitions, on the other hand, an entropy jump is always present at the transition point. Since $Q = T\Delta S$ is the latent heat of transition, this entropy jump is proportional to the latent heat of transition. Phase transitions with both positive and negative latent heats exist, i.e. entropy creation or reduction takes place in the system on supplying or withdrawing heat, but always $\Delta G = 0$ at the transition point. This does not violate the second law of thermodynamics since the system is not isolated thermally from the environment that may receive excess heat. This example is, of course, relevant to a living cell, if one were to speculate about jump-starting a living process by physical means or changing the state of a living system (Davies et al., 2011) from healthy to diseased (e.g. cancerous).

As emphasized earlier, a living cell constantly consumes energy to maintain its structure and vital functions. The energy comes basically in two forms: photons (in plants) and glucose-containing compounds (in animals). Glucose is easily utilized to synthesize ATP. Each glucose molecule gives rise to approximately $N = 30$ ATP molecules and the associated entropy production is given by (Daut, 1987)

$$\frac{dS}{dt} = \frac{\Delta G(\text{glucose})J(\text{ATP})}{NT}, \quad (17)$$

where $\Delta G(\text{glucose}) = 3 \times 10^6$ J/mol is the free energy of glucose oxidation and $J(\text{ATP}) = 10^{-13}$ mol/h is the flux of resultant ATP for a single cell (Kim et al., 1991). At the physiological temperature $T = 310$ K this results in an entropy rate of change for a single cell in the range of 10^{-14} J/Ks. This can be compared to only 0.7×10^{-17} J/Ks of entropy reduction due to DNA transmitted information, i.e. less than one thousandth as stated above. This is not surprising since many other processes are at work to keep the cell in its metastable (low entropy) state. First of all, the membrane itself consisting of phospholipids comprises some 60% of the cell's mass and presents a highly ordered structure requiring an entropy reduction to be put in place. Likewise, proteins and peptides are composed of up to several thousand atoms, each with fairly well specified positions, leading to a net entropy drop compared to a non-living state. Finally, approximately 50% of the metabolic energy of a cell is utilized in the process of ion pumping across the membrane (Rolfe and Brown, 1997), mainly as a result of trans-membrane potential and the work of ion pumps. The latter rely on molecular recognition mechanisms. These mechanisms, when a pump is activated, lower the entropy by binding the two molecules together. The subsequent placement of an ion or a macromolecule within the confines of a membrane permanently lowers the entropy by volume of exploration reduction, as discussed above. A release of a waste product into the environment results in a precisely opposite effect.

Chemical reactions may either absorb or release heat much like first order phase transitions, i.e. they may be either exothermic or endothermic. Since almost all living processes are in essence chains of chemical reactions, they can be analyzed from the point of view of entropy or information theory. In particular, enzymes use a fine-tuned selection mechanism of molecules that exhibit shape complementarity to a recognition pocket. This is called a lock-and-key mechanism by which enzymes force particular orientations of the catalytically reacting molecules and increase the corresponding reaction rates by several orders of magnitudes. Consequently, this process can be viewed as information processing whereby the shapes of binding domains are recognized, the molecules are optimally positioned for binding and in some cases particular bonds

are broken and others created. Some enzymes belonging to the class of allosteric proteins may adopt two or more stable conformations acting like switches, being activated in one conformation and inactive in some others. Following a binding and a catalysis event, enzymes return to their original conformation and act cyclically. From the point of view of information and entropy reduction, they do not overall decrease the entropy of the cell (Loewenstein, 1999). At best, they break even since any piece of information that an enzyme invests in a catalytic reaction is re-collected at the end of a cycle. Furthermore, it is important to stress, that the information necessary to perform a particular function (molecular recognition) is not entirely contained in an enzyme. In order for an enzyme to be effective, it must be activated by the environment in which it resides: cytoplasm containing inorganic ions and other molecules. Thus, one may say that the information is contained in the entire system, i.e. the cell. What then is the total information content of a cell?

6. Biological information

Production of DNA takes place even in non-replicating cells. A typical mammalian cell polymerizes approximately 2×10^8 nucleotides of DNA a minute into nuclear RNA (Brandhorst and McConkey, 1974), out of which only 5% end up in the cytoplasm coding for protein synthesis (Dreyfuss et al., 1993). Since there is redundancy in coding of nucleotide triplets for the 20 amino acids, the original 6 bits of information in DNA translate into $\log_2(20) = 4.2$ bits in a protein. Consequently, on the order of 0.7×10^6 bits/s are transmitted from the nucleus to the cytoplasm. This is augmented by a small fraction of information due to mitochondrial DNA (Alberts et al., 1994). As shown above, this is but a small fraction of the total information production (negative entropy) of a living cell. The vast majority of information is contained in the organized structure of the cell and its components.

Since the Shannon information formula employs probabilities of particular states, there are inherent dangers of incorrectly determining these probabilities, especially when this is done purely combinatorially as is for example the case in an amino acid or nucleic acid sequence analysis. However, this is not necessarily a random choice situation since different choices lead to different probability values (see Table 1), and a more appropriate description is given by the canonical ensemble Boltzmann distribution formula $p_i = p_0 \exp(-E_i/kT)$. In order to make this estimate work from first principles, one needs to know the energies E_i and hence the Hamiltonian for the system. Therefore, the apparent information estimate of $I = k_B \ln N^n$, where n is the number of members in a string, may be significantly larger than the true value of $-S$ from thermodynamic estimates of a given state, a maximum entropy state for equilibrium and hence a minimum information content. For a string of choices (e.g. amino acid sequence in a peptide or a nucleic acid sequence in a DNA or RNA molecule), this may lead to “basins of attraction” favoring some combinations more strongly over others. Furthermore, there could be evolutionary retention of favored choices and the establishment of hierarchies of order. An immense number has been defined as $I = 10^{110}$ and represents a clear computational barrier even from the point of view of cataloguing such an enormous number of objects. Immense numbers commonly appear in biology: both DNA and protein sequences are immense numbers arising from the sheer numbers of possible combinations in which these macromolecules may be formed. However, in view of the argument above, restricting the phase space by forming basins of attraction due to intramolecular interactions may result in a hugely reduced number of biologically relevant combinations one would encounter in practice.

Furthermore, a clear distinction between information and instruction should be made in the context of cell biology. While the former was introduced on purely statistical grounds as a measure of the number of choices possible when making a selection for a string of elements, instruction implies the existence of a message, a messenger, and a reader who would be able execute the message. A classic example of this is the synthesis of amino acids contained in the triples of DNA and RNA base pairs. While there is the same information content in every triplet, namely $k_B \ln 4^3 = 6$ bits, some amino acids are coded uniquely by a single triplet and some by two, three, four, etc., different codons (see Table 1). This is obvious in view of the fact that there are 64 possible triplets of base pairs while only 20 distinct amino acids, hence the redundancy. A similar difference between information and instruction can be found in the genome where in addition to the coding sequences of DNA, some of which are of vital importance to the very survival of a given organism, one finds so-called junk DNA that has apparently no coding value but represents a vast majority of the DNA sequence. We stress here that information can be easily confused with instruction.

DNA and RNA are thought to be such biological messengers, and so are hormones and various signaling molecules such as kinase and phosphatase enzymes. However, as shown earlier, it appears that a vast majority of information content is not instructional in nature. This is akin to simple algorithms like the logistic map or fractal recursive relations that give rise to great mathematical complexity of the results that follow. Similarly, DNA can be viewed as an algorithm that spans an awe-inspiring complexity of living cells. While coding for protein synthesis is contained in the genetic code, it is most improbable that details of structure formation need special coding. They most likely unfold due to self-organization inherent in the dynamics of the synthesized products. This type of behavior is well-known to both physicists and chemists.

In non-equilibrium systems such as autocatalytic chemical reactions of the Brusselator type (Prigogine, 1980), order is created and sustained by means of non-linear interactions and external forcing. Another important property of non-linear systems is the possibility of self-assembly, for example in pattern forming crystal growth. This provides an example where there is no necessity for an instruction-driven creation of order and structure. Sometimes, when discussing the assembly of bio-matter, concern is unduly given to the need for instruction in putting the building blocks of matter together piece by piece. While there are clear instructions for the amino acid sequences in the genome, the details of higher order structure formation need no special encoding. They may emerge spontaneously as an attractor in non-linear dynamical system that we call a living cell as a result of biological self-organization (Kauffmann, 1993).

We postulate, therefore, the existence of two types of information in biological systems: (a) structural information – i.e. negative entropy – and (b) functional information. The former is simply related to a neat and tight packing of the various molecules into macromolecules and macromolecules into organelles that comprise the cell. The latter on the other hand, pertains to the functioning of the cell and hence the rate and amount of chemical reactions taking place. The two forms are somewhat related but not identical. Imagine the construction of a car as an analogy. It may look perfectly good but if the gas line is cut, it will not be able to run. The same holds true for a living cell. Some key reactions if not properly executed, will lead to the cell death. While structural information should be maximized meaning entropy reduction by the cell, functional information is concerned with how rapidly information is being exchanged. Moreover, the cell must make sure that the information exchange between its parts is carried out faithfully, i.e. it is error-free. In most sensitive areas such as the coding regions of DNA, error correction mechanisms are actively involved by the

use of DNA repair enzymes whose work assures proper information transfer but it requires energy expense on the part of the cell.

There has been significant progress in our quantitative understanding of information content involved in DNA, RNA championed by Schneider (2001) and transcription factors by Kim et al. (2003) and by for macromolecules Sadovsky (2003) and even whole microbial genomes by Lio (2003).

7. Conclusions

The various points of reference regarding the nature of the living state undoubtedly reflect the prevailing Zeitgeist of the period in which a given theory has been created. The viewpoint of representing the cell as a machine, or even a factory, closely mirrors the worldview of the industrial revolution of the 19th century. Likewise, the currently popular opinion that living cells are intensely engaged in some type of computation is closely linked with the information technology revolution ushered into the second half of the 20th century as a result of the proliferation of computers in our daily lives. Both points of view have merits, i.e. the cell obeys the laws of physics such as the first law of thermodynamics and hence can be viewed as a thermodynamic machine and simultaneously it locally acts against the second law of thermodynamics by creating structural and functional order. In other words, it creates and maintains information by expending energy produced from nutrient in the form of ATP and GTP molecules. The latter aspect is thermodynamically analogous to the way a refrigerator works. However, a biological cell also processes information and engages in signaling thereby actively performing computation. It is safe to say that living cells can be viewed as both micro-factories (with nano-machines performing individual tasks), and biological computers whose nano-chips are the various proteins and peptides in addition to DNA and RNA. Most of the cell is what we could call hardware while a small fraction is analogous to computer software (for example the genetic code in the DNA that instructs for the synthesis of proteins).

Finally, it is worth stating that the role of information and entropy in living cells, which has been the focus of this article, is becoming a topic of serious investigations in connection with health and disease. For example, this is becoming increasingly clear in cancer where molecular changes can be quantified in terms of entropy gain or information loss (Frieden and Gatenby, 2011; Davies et al., 2012). In the present issue, entropic changes can be seen affecting embryogenesis at the level of intracellular re-organization of the cytoskeleton, directly seen in the Ising-like phase transition of the cortical microtubules (Tuszynski and Gordon, 2012) where their interactions overcome entropic disorder. This has been successfully simulated in a computational model of microtubule dynamics (Nouri et al., 2012).

Acknowledgments

Research for this project has been supported by an NSERC (Canada) grant awarded to JAT. This work was supported in part by NIH grant U54 CA143682 awarded to PCWD.

References

- Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., Watson, J.D., 1994. *Molecular Biology of the Cell*. Garland Publishing, New York.
- Boltzmann, L., 1872. Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. *Wiener Berichte* 66, 275–370.
- Brandhorst, B.P., McConkey, E.H., 1974. Stability of nuclear RNA in mammalian cell. *J. Mol. Biol.* 85, 451–463.
- Clausius, R., 1865. Über die Wärmeleitung gasförmiger Körper. *Ann. Phys.* 125, 353–400.
- Cook, N.D., 1984. The transmission of information in natural systems. *J. Theor. Biol.* 108, 349–367.
- Daut, J., 1987. The living cell as an energy-transducing machine. *Biochem. Biophys. Acta* 895, 41–62.
- Davies, P.C., Demetrius, L., Tuszynski, J.A., 2011. Cancer as a dynamical phase transition. *Theor. Biol. Med. Model.* 8, 30.
- Davies, P.C., Demetrius, L., Tuszynski, J.A., 2012. Implications of quantum metabolism and natural selection for the origin of cancer cells and tumor progression. *AlP Adv.*, Special Issue Phys. Cancer 2, 011101.
- Dewey, T.G., 1996. Algorithmic complexity of a protein. *Phys. Rev. E: Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.* 54, R39–R41.
- Dewey, T.G., Strait, B.J., 1996. Multifractals, decoded walks and the ergodicity of protein sequences. In: Hunter, L., Klein, T.E. (Eds.), *Pacific Symposium on Bio-computing*. World Scientific, Singapore, pp. 216–229.
- Dreyfuss, G., Matunis, M.J., Piñol-Roma, S., Burd, C.G., 1993. hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.* 62, 289–321.
- Driesch, H.A.E., Morgan, T.H., 1895. Zur Analysis der ersten Entwicklungsstadien des Ctenophoreneies. II. Von der Entwicklung ungeführter Eier mit Protoplasma-defektes (On the analysis of the first development stages of the Ctenophore egg. II. Of the development of uncleaved eggs with protoplasm defects). *Arch. Entwicklungsmech. Org.* 2, 216–224.
- Frieden, B.R., Gatenby, R.A., 2011. Information dynamics in living systems: prokaryotes, eukaryotes, and cancer. *PLoS ONE* 6, e22085.
- Gibbs, J.W., 1902. *Elementary Principles in Statistical Mechanics: Developed With Especial Reference to the Rational Foundation of Thermodynamics*. C. Scribner, New York.
- Gilbert, E.N., 1966. Information theory after 18 years. *Science* 152, 320–326.
- Goel, N.S., Campbell, R.D., Gordon, R., Rosen, R., Martinez, H., Ycas, M., 1970. Self-sorting of isotropic cells. *J. Theor. Biol.* 28, 423–468.
- Goel, N.S., Campbell, R.D., Gordon, R., Rosen, R., Martinez, H., Ycas, M., 1975. Self-sorting of isotropic cells. In: Mostow, G.D. (Ed.), *Mathematical Models for Cell Rearrangement*. Yale University Press, New Haven, pp. 100–144.
- Gordon, R., 1999. *The Hierarchical Genome and Differentiation Waves: Novel Unification of Development, Genetics and Evolution*. World Scientific/Imperial College Press, Singapore/London.
- Gordon, R., Goel, N.S., Steinberg, M.S., Wiseman, L.L., 1972. A rheological mechanism sufficient to explain the kinetics of cell sorting. *J. Theor. Biol.* 37, 43–73.
- Gordon, R., Goel, N.S., Steinberg, M.S., Wiseman, L.L., 1975. A rheological mechanism sufficient to explain the kinetics of cell sorting. In: Mostow, G.D. (Ed.), *Mathematical Models for Cell Rearrangement*. Yale University Press, New Haven, pp. 196–230.
- Houliston, E., Carré, D., Johnston, J.A., Sartet, C., 1993. Axis establishment and microtubule-mediated waves prior to first cleavage in *Beroe ovata*. *Development* 117, 75–87.
- Johnson, H.A., 1970. Information theory in biology after 18 years. *Science* 168, 1545–1550.
- Kauffman, S.A., 1993. *The Origins of Order: Self-organization and Selection in Evolution*. Oxford University Press, New York.
- Kim, H.D., Koury, M.J., Lee, S.J., Im, J.H., Sawyer, S.T., 1991. Metabolic adaptation during erythropoietin-mediated differentiation of mouse erythroid cells. *Blood* 77, 387–392.
- Kim, J.T., Martinetz, T., Polani, D., 2003. Bioinformatic principles underlying the information content of transcription factor binding sites? *J. Theor. Biol.* 220, 529–544.
- Kinoshita, M., 2009. Importance of translational entropy of water in biological self-assembly processes like protein folding. *Int. J. Mol. Sci.* 10, 1064–1080.
- Lambert, G.R., 1984. Enzymic editing mechanisms and the origin of biological information transfer. *J. Theor. Biol.* 107, 387–403.
- Landau, L.D., Lifshitz, E.M., 1969. *Statistical Physics*. Addison-Wesley, New York.
- Lio, P., 2003. Statistical bioinformatic methods in microbial genome analysis. *Bioessays* 25, 266–273.
- Loewenstein, W.R., 1999. *The Touchstone of Life*. Oxford University Press, Oxford.
- Marina, D., Martín, M., Sabater, B., 2009. Entropy decrease associated to solute compartmentalization in the cell. *Biosystems* 98, 31–36.
- Maturana, H.R., Varela, F.J., 1980. *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel, Dordrecht.
- Mawell, J.C., 1871. Remarks on the mathematical classification of physical quantities. *Proc. Lond. Math. Soc.* s1–3, 224–233.
- Morowitz, H.J., 1955. Some order–disorder considerations in living systems. *Bull. Math. Biophys.* 17, 81–86.
- Nouri, C., Tuszynski, J.A., Wiebe, M., Gordon, R., 2012. Simulation of the effects of microtubules in the cortical rotation of amphibian embryos in normal and zero gravity. *Biosystems* 109, 444–449.
- Pande, V.S., Grosberg, A.Y., Tanaka, T., 1994. Nonrandomness in protein sequences: evidence for a physically driven stage of evolution? *Proc. Natl. Acad. Sci. U.S.A.* 91, 12972–12975.
- Penrose, O., 1979. Foundations of statistical mechanics. *Rep. Prog. Phys.* 42, 1937–2006.
- Portet, S., Tuszynski, J.A., Dixon, J.M., Sataric, M.V., 2003. Models of spatial and orientational self-organization of microtubules under the influence of gravitational fields. *Phys. Rev. E: Stat. Nonlin. Soft Matter Phys.* 68 (2 Pt 1), 021903.
- Prigogine, I., 1980. *From Being to Becoming: Time and Complexity in the Physical Sciences*. W.H. Freeman and Co., San Francisco.
- Rappaport, R., 1996. *Cytokinesis in Animal Cells*. Cambridge University Press, New York.
- Rolfe, D.F., Brown, G.C., 1997. Cellular energy utilization and molecular origin of standard metabolic rate in mammals. *Physiol. Rev.* 77, 731–758.

- Rosen, R., 1991. *Life Itself: A Comprehensive Inquiry into the Nature, Origin and Fabrication of Life*. Columbia University Press, New York.
- Sadovsky, M.G., 2003. Comparison of real frequencies of strings vs. the expected ones reveals the information capacity of macromolecules. *J. Biol. Phys.* 29, 23–38.
- Saenger, W., 1984. *Principles of Nucleic Acid Structure*. Springer-Verlag, New York.
- 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.
- Schrödinger, E., 1967. *What is Life? The Physical Aspect of the Living Cell*. Cambridge University Press, Cambridge.
- Shannon, C.E., 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27 (379–423), 623–656.
- Shen, S., Kai, B., Ruan, J., Huzil, J.T., Carpenter, E., Tuszynski, J.A., 2006. Probabilistic analysis of the frequencies of amino acid pairs within characterized protein sequences. *Physica A* 370, 651–662.
- Sinden, R., 1994. *DNA Structure and Function*. Academic Press, San Diego.
- Speksnijder, J.E., Sardet, C., Jaffe, L.F., 1990. Periodic calcium waves cross ascidian eggs after fertilization. *Dev. Biol.* 142, 246–249.
- Steinberg, M.S., 1996. Adhesion in development: an historical overview. *Dev. Biol.* 180, 377–388.
- Strait, B.J., Dewey, T.G., 1996. Shannon information entropy of protein sequences. *Biophys. J.* 71, 148–155.
- Stryer, L., 1981. *Biochemistry*. W.H. Freeman and Co., San Francisco.
- Szilard, L., 1929. On the decrease of entropy in a thermodynamic state by the intervention of intelligent beings. *Z. Phys.* 53, 840–856.
- Trinkaus, J.P., 1984. *Cells into Organs: The Forces That Shape the Embryo*, 2nd ed. Prentice-Hall, Englewood Cliffs, New Jersey.
- Tuszynski, J.A., Gordon, R., 2012. A mean field Ising model for cortical rotation in amphibian one-cell stage embryos. *Biosystems* 109, 381–389.
- Yoneda, M., Kobayakawa, Y., Kubota, H.Y., Sakai, M., 1982. Surface contraction waves in amphibian eggs. *J. Cell Sci.* 54, 35–46.