Version 4

A Rate Distortion approach to protein symmetry

Rodrick Wallace, Ph.D.
Division of Epidemiology
The New York State Psychiatric Institute*

March 14, 2010

Abstract

A spontaneous symmetry breaking argument is applied to the problem of protein form, via a Rate Distortion analysis of the relation between genome coding and the final condensation of the protein ‘molten globule’. The Rate Distortion Function, under coding constraints, serves as a temperature analog, so that low values act to drive proteins to simple symmetries. The Rate Distortion Function itself is significantly constrained by the availability of metabolic free energy. This work extends Thusty’s (2007) elegant exploration of the evolution of the genetic code, suggesting that rate distortion considerations may play a critical role across a broad spectrum of molecular expressions of evolutionary process.

Key Words groupoid, information theory, phase transition, protein symmetry, Rate Distortion, spontaneous symmetry breaking

1 Introduction

The symmetries of folded proteins remain something of a scientific mystery, in spite of decades of concerted attention. Figure 1, from Goodsell and Olson (2000), shows a few typical examples. More general, but less overtly ‘symmetric’, conformations, however, involve finite tilings of helices, sheets, and attachment loops that would seem better described using groupoid methods, following the arguments of Weinstein (1996): As Wolynes (1996) put the matter, “It is the inexact symmetries of biological molecules that are most striking”.

Anfinsen’s (1973) thermodynamic hypothesis has strongly dominated most thinking on the subject: the native state of a protein has the lowest Gibbs free

*Address correspondence to Rodrick Wallace, 549 W. 123 St., Apt. 16F, New York, NY, 10027 USA, email wallace@pi.cpmc.columbia.edu.

Figure 1: From Godsell and Olson, (2000). Crystallographic point group symmetries. Proteins with each of the crystallographic point group symmetries have been found. Point group symbols are included below each protein structure (e.g., C1 and D2), and the number of identical subunits in each group is included below and to the right of the structure. One subunit is shaded in each example. Other, noncrystallographic, point groups are consistent with the entantiomorphic nature of proteins, including complicated cyclic and dihedral symmetries. The more ubiquitous tilings of helices, sheets, and loops can probably be best described using groupoids, most directly seen as disjoint unions of groups, for which a 'product' is only locally defined (e.g., Weinstein, 1996).

energy, determined by the interaction of the amino acid sequence and the embedding environment. More recent work (e.g., summarized in Lei and Huang, 2010) suggests that large, complex proteins may have native configurations representing kinetically accessible, rather than thermodynamically minimal, states. Andre et al. (2008) explore the central insight that “...selection is only likely to operate on primordial complexes with sufficient initial interaction energy to at least partially overcome the entropic costs of association of the monomers; evolution can only optimize a complex that is populated sufficiently to confer a benefit on the organism”.

Here we will attempt to finesse this perspective by invoking, first, a rate distortion argument applied to the transmitted signal represented by the translation of the genome into the final condensation of the molten globule of the resulting amino acid string. The second stage is to examine the likely effect of limited availability of metabolic free energy, or processing time, on the fidelity of this transmission. The observation of a possible ‘metabolic singularity’ limiting that fidelity suggests, in turn, that possible protein symmetries were, in some measure, limited by energy availability in prebiotic circumstances, leading to early evolutionary lock-in by path dependence.

The argument, an extension of Tlusty’s (2007) insights regarding the role of rate distortion constraints in evolutionary process, seems fairly direct, based on standard material from statistical physics and information theory.

2 Spontaneous symmetry breaking

Landau’s theory of phase transitions (Landau and Lifshitz, 2007) assumes that the free energy of a system near criticality can be expanded in a power series of some ‘order parameter’ ϕ representing a fundamental measurable quantity, that is, a symmetry invariant. One writes

$$F_0 = \sum_{k=m}^{p(>m)} A_k \phi^k, \quad (1)$$

with $A_2 \approx \alpha(T - T_c)$ sufficiently close to the critical temperature T_c . This mean field approach can be used to describe a variety of second-order effects for $p = 4$ or $p = 6$, $A_3 = 0$ and $A_4 > 0$, and first order phase transitions (requiring latent heat) for either $p = 6$, $A_3 = 0$, $A_4 < 0$ or $p = 4$ and $A_3 \neq 0$. These can be both temperature induced (for $m = 2$) and field induced (for $m = 1$).

Minimization of F_0 with respect to the order parameter yields the average value of ϕ , $\langle \phi \rangle$, which is zero above the critical temperature and non-zero below it. In the absence of external fields, the second-order transition occurs at $T = T_c$, while the first-order, needing latent heat, occurs at $T_c^* = T_c + A_4^2/4\alpha A_6$. In the latter case thermal hysteresis arises between $T_s \equiv T_c + A_4^2/3\alpha A_6$ and T_c .

A more accurate approximation involves an expression that recognizes the effect of coarse-graining, adding a term in $\nabla^2\phi$ and integrating over space rather than summing. Regimes dominated by this gradient will show behaviors analogous to those described using the one dimensional Landau-Ginzburg equation, which, among other things, characterizes superconductivity.

The Landau formalism quickly enters deep topological waters (Pettini, 2007, pp. 42-43; Landau and Lifshitz, 2007, pp. 459-466). The essence of Landau's insight was that phase transitions without latent heat – second order transitions – were usually in the context of a significant symmetry change in the physical states of a system, with one phase, at higher temperature, being far more symmetric than the other. A symmetry is lost in the transition, a phenomenon called spontaneous symmetry breaking. The greatest possible set of symmetries in a physical system is that of the Hamiltonian describing its energy states. Usually states accessible at lower temperatures will lack symmetries available at higher temperatures, so that the lower temperature phase is the less symmetric: The randomization of higher temperatures ensures that higher symmetry/energy states will then be accessible to the system.

At the lower temperature an order parameter must be introduced to describe the system's physical states – some extensive quantity like magnetization. The order parameter will vanish at higher temperatures, involving more symmetric states, and will be different from zero in the less symmetric lower temperature phase.

This can be formalized, following Pettini (2007), as follows. Consider a thermodynamic system having a free energy F which is a function of temperature T , pressure P , and some other extensive macroscopic parameters m_i , so that $F = F(P, T, m_i)$. The m_i all vanish in the most symmetric phase, so that, as a function of the m_i , $F(P, T, m_i)$ is invariant with respect to the transformations of the symmetry group G_0 of the most symmetric phase of the system when all $m_i \equiv 0$.

The state of the system can be represented by a vector $|m\rangle = |m_1, \dots, m_n\rangle$ in a vector space $\mathcal{E}$. Now, within $\mathcal{E}$, construct a linear representation of the group G_0 that associates with any $g \in G_0$ a matrix $M(g)$ having rank n . In general, the representation $M(g)$ is reducible, and we can decompose $\mathcal{E}$ into invariant irreducible subspaces $\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_k$, having basis vectors $|e_i^{(n)}\rangle$ with $n = 1, 2, \dots, n_i$ and $n_i = \dim \mathcal{E}_i$. The state variables m_i are transformed into new variables $\eta_i^{(n)} = \langle e_i^{(n)} | m \rangle$, where the bracket represents an inner product.

In terms of irreducible representations $D_i(g)$ induced by $M(g)$ in $\mathcal{E}_i$ we have

$$M(g) = D_1(g) \oplus D_2(g) \oplus \dots \oplus D_k(g).$$

If at least one of the $\eta_i^{(n)}$ is nonzero, then the system no longer has the symmetry G_0 . This symmetry has been broken, and the new symmetry group is G_i , associated with the representation $D_i(g)$ in $\mathcal{E}_i$. The variables $\eta_i^{(n)}$ are the new order parameters, and the free energy is now $F = F(P, T, \eta_i^{(n)})$. For a physical system the actual values of the η as functions of P and T can be variationally determined by minimizing the free energy F .

Two essential features distinguish information systems, like the translation of a genome into a folded protein, from this simple physical model.

First, order parameters cannot always be determined by simple minimization procedures, as ‘cognitive’ information systems can, within their contextual constraints (which particularly include available metabolic free energy), choose states which are not energy or other extrema (e.g., Wallace and Wallace, 2008).

Second, the essential symmetry of information sources is quite often driven by groupoid, rather than group, structures (e.g., Wallace, 2010). One must then engage the full transitive orbit/isotropy group decomposition, and examine groupoid representations (e.g., Bos, 2007; Buneci, 2003) configured about the irreducible representations of the isotropy groups. This observation seems particularly relevant given the usual helix/sheet/connecting loop tilings that characterize most elaborate protein conformations (Wolynes, 1996).

3 A little information theory

Here we think of the machinery listing a sequence of codons as communicating with machinery that produces amino acids, folds them in context, and produces the final symmetric protein. We suppose it possible to compare what is actually produced with what should have been produced, perhaps by a simple survival of the fittest selection mechanism, perhaps via some more sophisticated error-correcting systems.

Suppose a sequence of signals is generated by a biological information source Y having output $y^n = y_1, y_2, \dots$ – codons. This is ‘digitized’ in terms of the observed behavior of the system with which it communicates, say a sequence of ‘observed behaviors’ $b^n = b_1, b_2, \dots$ – amino acids and their folded protein structure. Assume each b^n is then deterministically retranslated back into a reproduction of the original biological signal, $b^n \rightarrow \hat{y}^n = \hat{y}_1, \hat{y}_2, \dots$.

Define a distortion measure $d(y, \hat{y})$ which compares the original to the retranslated path. Many distortion measures are possible. The Hamming distortion is defined simply as

$$d(y, \hat{y}) = 1, y \neq \hat{y}$$

$$d(y, \hat{y}) = 0, y = \hat{y}.$$

For continuous variates the squared error distortion is just $d(y, \hat{y}) = (y - \hat{y})^2$.

There are many such possibilities. The distortion between *paths* y^n and $\hat{y}^n$ is defined as $d(y^n, \hat{y}^n) \equiv \frac{1}{n} \sum_{j=1}^n d(y_j, \hat{y}_j)$.

A remarkable fact of the Rate Distortion Theorem is that *the basic result is independent of the exact distortion measure chosen* (Cover and Thomas, 1991; Dembo and Zeitouni, 1998).

Suppose that with each path y^n and b^n -path retranslation into the y -language, denoted $\hat{y}^n$, there are associated individual, joint, and conditional probability distributions $p(y^n), p(\hat{y}^n), p(y^n, \hat{y}^n), p(y^n | \hat{y}^n)$.

The average distortion is defined as

$$(2) \quad D \equiv \sum_{y^n} p(y^n) d(y^n, \hat{y}^n).$$

It is possible, using the distributions given above, to define the information transmitted from the Y to the $\hat{Y}$ process using the Shannon source uncertainty of the strings:

$$I(Y, \hat{Y}) \equiv H(Y) - H(Y|\hat{Y}) = H(Y) + H(\hat{Y}) - H(Y, \hat{Y}),$$

where $H(...)$ is the standard joint, and $H(...|...)$ the conditional, Shannon uncertainties (Cover and Thomas, 1991; Ash, 1990).

If there is no uncertainty in Y given the retranslation $\hat{Y}$, then no information is lost, and the systems are in perfect synchrony.

In general, of course, this will not be true.

The *rate distortion function* $R(D)$ for a source Y with a distortion measure $d(y, \hat{y})$ is defined as

$$(3) \quad R(D) = \min_{p(y, \hat{y}); \sum_{(y, \hat{y})} p(y) p(y|\hat{y}) d(y, \hat{y}) \leq D} I(Y, \hat{Y}).$$

The minimization is over all conditional distributions $p(y|\hat{y})$ for which the joint distribution $p(y, \hat{y}) = p(y)p(y|\hat{y})$ satisfies the average distortion constraint (i.e., average distortion $\leq D$).

The *Rate Distortion Theorem* states that $R(D)$ is the minimum necessary rate of information transmission which ensures the communication between the biological vesicles does not exceed average distortion D . Thus $R(D)$ defines a minimum necessary channel capacity. Cover and Thomas (1991) or Dembo and Zeitouni (1998) provide details. The rate distortion function has been calculated for a number of systems.

We reiterate an absolutely central fact characterizing the rate distortion function: Cover and Thomas (1991, Lemma 13.4.1) show that $R(D)$ is necessarily a decreasing convex function of D for any reasonable definition of distortion.

That is, $R(D)$ is always a reverse J-shaped curve. This will prove crucial for the overall argument. Indeed, convexity is an exceedingly powerful mathematical condition, and permits deep inference (e.g., Rockafellar, 1970). Ellis (1985, Ch. VI) applies convexity theory to conventional statistical mechanics.

For a Gaussian channel having noise with zero mean and variance σ^2 (Cover and Thomas, 1991),

$$\begin{aligned}
R(D) &= 1/2 \log[\sigma^2/D], 0 \leq D \leq \sigma^2 \\
R(D) &= 0, D > \sigma^2.
\end{aligned}
\tag{4}$$

Recall, now, the relation between information source uncertainty and channel capacity (e.g., Ash, 1990):

$$H[X] \leq C,$$

(5)

where H is the uncertainty of the source X and C the channel capacity, defined according to the relation (Ash, 1990)

$$C \equiv \max_{P(X)} I(X|Y),$$

(6)

where $P(X)$ is chosen so as to maximize the rate of information transmission along a channel Y .

Note that for a parallel set of noninteracting channels, the overall channel capacity is the sum of the individual capacities, providing a powerful ‘consensus average’ that does not apply in the case of modern molecular coding.

Finally, recall the analogous definition of the rate distortion function above, again an extremum over a probability distribution.

Our own work (Wallace and Wallace, 2008) focuses on the homology between information source uncertainty and free energy density. More formally, if $N(n)$ is the number of high probability ‘meaningful’ – that is, grammatical and syntactical – sequences of length n emitted by an information source X , then, according to the Shannon-McMillan Theorem, the zero-error limit of the Rate Distortion Theorem (Ash, 1990; Cover and Thomas, 1991; Khinchin, 1957),

$$\begin{aligned}
H[X] &= \lim_{n \rightarrow \infty} \frac{\log[N(n)]}{n} \\
&= \lim_{n \rightarrow \infty} H(X_n | X_0, \dots, X_{n-1})
\end{aligned}$$

$$= \lim_{n \rightarrow \infty} \frac{H(X_0, \dots, X_n)}{n+1},$$

(7)

where, again, $H(\dots|\dots)$ is the conditional and $H(\dots, \dots)$ is the joint Shannon uncertainty.

In the limit of large n , $H[X]$ becomes homologous to the free energy density of a physical system at the thermodynamic limit of infinite volume. More explicitly, the free energy density of a physical system having volume V and partition function $Z(\beta)$ derived from the system's Hamiltonian – the energy function – at inverse temperature β is (e.g., Landau and Lifshitz 2007)

$$F[K] = \lim_{V \rightarrow \infty} -\frac{1}{\beta} \frac{\log[Z(\beta, V)]}{V} \equiv$$

$$\lim_{V \rightarrow \infty} \frac{\log[\hat{Z}(\beta, V)]}{V},$$

with $\hat{Z} = Z^{-1/\beta}$. The latter expression is formally similar to the first part of equation (7), a circumstance having deep implications: Feynman (2000) describes in great detail how information and free energy have an inherent duality. Feynman, in fact, defines information precisely as the free energy needed to erase a message. The argument is surprisingly direct (e.g., Bennett, 1988), and for very simple systems it is easy to design a small (idealized) machine that turns the information within a message directly into usable work – free energy. Information is a form of free energy and the construction and transmission of information within living things consumes metabolic free energy, with nearly inevitable losses via the second law of thermodynamics. If there are limits on available metabolic free energy there will necessarily be limits on the ability of living things to process information.

Figure 2 presents a schematic of the mechanism: As the complexity of a dynamic physiological information process rises, that is, as H increases, its free energy content increases linearly. The metabolic free energy needed to construct and maintain the physiological systems that instantiate H should, however, be expected to increase nonlinearly with it, hence the ‘translation gap’ of the figure. Below we will infer a kind of nonequilibrium thermodynamics driven by gradients in an entropy-like factor constructed from H , analogs to empirical Onsager relations. Figure 2 suggests that H may indeed be a good, if highly nonlinear, index of large-scale free energy dynamics.

Conversely, information source uncertainty has an important heuristic interpretation that Ash (1990) describes as follows:

Figure 2: Nonlinear increase in metabolic free energy needed to maintain and generate linear increase in the information source uncertainty of a complex physiological process. H is seen to ‘leverage’ metabolic expenditures, parameterizing a more complicated nonequilibrium thermodynamics.

[W]e may regard a portion of text in a particular language as being produced by an information source. The probabilities $P[X_n = a_n | X_0 = a_0, \dots, X_{n-1} = a_{n-1}]$ may be estimated from the available data about the language; in this way we can estimate the uncertainty associated with the language. A large uncertainty means, by the [Shannon-McMillan Theorem], a large number of ‘meaningful’ sequences. Thus given two languages with uncertainties H_1 and H_2 respectively, if $H_1 > H_2$, then in the absence of noise it is easier to communicate in the first language; more can be said in the same amount of time. On the other hand, it will be easier to reconstruct a scrambled portion of text in the second language, since fewer of the possible sequences of length n are meaningful.

In sum, if a biological system characterized by H_1 has a richer and more complicated internal communication structure than one characterized by H_2 , then necessarily $H_1 > H_2$ and system 1 represents a more energetic process than system 2, and by the arguments of figure 2, may trigger even greater metabolic free energy dynamics.

By equations (5), (6), and (7), the Rate Distortion Function, $R(D)$ is likewise a free energy measure, constrained by the availability of metabolic free energy.

4 Mixing metaphors

Ash's comment leads directly to a model in which the average distortion between the initial codon stream and the final form of the folded amino acid stream, the protein, becomes a dominant force, particularly in an evolutionary context in which fidelity of codon expression has survival value. The direct model examines the distortion between the codon stream and the folded protein structure, via the free energy of the possible folded conformations.

Suppose there are n possible folding schemes. The simplest approach is to assume that, for a given rate distortion function and distortion measure, $R(D)$, under evolutionary selection constraints, serves much as an external temperature bath for the possible distribution of conformation free energies, the set $\{\mathcal{H}_1, \dots, \mathcal{H}_n\}$. That is, low distortion, represented by a high rate of transmission of information between codon machine and amino acid/protein folding machine, permits more complicated symmetries according to the classic formula

$$Pr[\mathcal{H}_j] = \frac{\exp[-\mathcal{H}_j/\lambda R(D)]}{\sum_{i=1}^n \exp[-\mathcal{H}_i/\lambda R(D)]}, \quad (8)$$

where $Pr[\mathcal{H}_j]$ is the probability of folding scheme j having conformational free energy $\mathcal{H}_j$.

We are, in essence, assuming that $Pr[\mathcal{H}_j]$ is a one parameter distribution in the 'extensive' quantity $R(D)$ (monotonic convex, however, in D) rather than a simple 'intensive' temperature-analog. This is permitted under the 'structurally intensive' circumstance of codon stream translation into amino acid stream.

The free energy Morse Function associated with this probability is

$$F_R = -\lambda R(D) \log[\sum_{i=1}^n \exp[-\mathcal{H}_i/\lambda R(D)]]. \quad (9)$$

Applying a spontaneous symmetry lifting argument to F_R generates topological transitions in folded protein structure as the 'temperature' $R(D)$ increases, i.e., as the average distortion D declines, via the inherent convexity of the Rate Distortion Function. That is, as the channel capacity connecting codon machines with amino acid/protein folding machines increases, more complex conformations become possible.

What, then, drives $R(D)$?

5 Metabolic singularity

A probability density function for $R(D)$ at a given intensive index of embedding metabolic energy, M , can most simply be described using the same approach as equation (8):

$$Pr[R(D), \kappa M] = \frac{\exp[-R(D)/\kappa M]}{\int_{D_{min}}^{D_{max}} \exp[-R(D)/\kappa M] dD}, \quad (10)$$

where κM represents the synergism between the intensity and physiological availability of the embedding metabolic free energy. Below we will extend the argument to include characteristic development time.

At a fixed κM , the mean of R is

$$\langle R \rangle = \int_{D_{min}}^{D_{max}} R(D) P[D, \kappa M] dD. \quad (11)$$

For the Gaussian channel, with $R = (1/2) \log(\sigma^2/D)$; $0 \leq D \leq \sigma^2$,

$$\langle R \rangle = \kappa M / (1 + 2\kappa M). \quad (12)$$

Note that the variance σ^2 has disappeared, so that $\langle R \rangle$ represents not only the mean with respect to distortion, but is also a kind of consensus average, albeit one without the benefit of parallelism.

If there is a one parameter distribution of free energy input intensity measures characterized, say, by some ‘temperature-equivalent’ T_e , then, crudely, M can be replaced by its average

$$\langle M \rangle = \frac{\int_0^\infty \exp[-M/\beta T_e] dM}{\int_0^\infty \exp[-M/\beta T_e] dM} = \beta T_e. \quad (13)$$

Figure 3: The Metabolic Singularity: For a single channel, large values of the average rate distortion function, $\langle R \rangle$, and hence low distortion in codon translation to final protein conformation, require explosively high rates of physiologically available metabolic free energy beyond a certain point.

Note that equations (12) and (13) can be used to reexpress equation (10) in terms of the ‘intensive’ quantity T_e , i.e., replacing $R(D)$ with $\kappa\beta T_e / (1 + 2\kappa\beta T_e)$.

The rapid topping-out of the rate distortion function with κM (or $\kappa\beta T_e$) in the toy model of equation (12) may be an example of the dynamic suggested by figure 2. While small changes in κM can have linear impact early on, their effect rapidly becomes muted, suggesting the possibility that evolutionary matters regarding accurate transcription and protein folding were settled well before the aerobic transition: Figure 3 shows equation (12) solved for κM in terms of $\langle R \rangle$. The synergism of metabolic energy with the efficiency of its use that is needed to actually produce an increase in the average Rate Distortion Function – and hence to reduce the distortion-measure translation/folding error – literally explodes, creating a serious barrier to accurate translation, beyond a certain level that may have been reached early in life’s history on Earth.

In neural systems (as in current multi-core computer chips), the solution to this conundrum is to permit multiple parallel channels that simply are not available to the protein coding scheme.

This suggests the necessity of replacing κM with a product $\gamma\tau M$ where τ

is a characteristic allowed reaction time. If τ and M are constrained, then, according to these arguments, protein symmetry possibilities will necessarily be constrained as well.

If we follow arguments similar to Section 6 of Wallace (2010), we can take $\gamma\tau M$ as a product of eigenvalues, the determinant of a Hessian matrix representing a Morse Function f on some underlying background manifold, $\mathcal{M}$, characterized in terms of unspecified embedding ecosystem variables $\mathcal{X} = \{x^1, \dots, x^n\}$. Let $h_{i,j} = \partial^2 f / \partial x^i \partial x^j$ and $\gamma\tau M = \det(h_{i,j})$. If $\gamma\tau M$ is a volume element, $h_{i,j}$ has only positive eigenvalues, and f has index zero, so that $\mathcal{M}$ is a simply connected geometric object. A number of simply connected spaces can be patched together to create more complicated topological structures subject to punctuated ‘ecosystem resilience’ transitions that we will not explore further here.

6 Discussion and conclusions

In this model, the fidelity of the translation between genome and final protein conformation, measured by a ‘structurally intensive’ rate distortion function, serves as an evolutionally-sculpted temperature analog to determine the possible phase transitions defining different degrees of protein symmetry. These symmetries may be characterized by finite groupoid tilings as well as by the kinds of structures shown in figure 1. Fidelity of replication is, in turn, constrained by the availability of metabolic free energy, either in terms of energy intensity or allowed development time.

Following the arguments of Wallace (2009, 2010), it is not difficult to construct evolutionary ‘just so’ stories in which early prebiotic limitations on metabolic free energy sources (e.g., Canfield et al., 2006) would limit possible protein symmetries, ultimately locking-in a set of simply symmetric proteins via evolutionary path dependence. Subsequent availability of more energetic metabolic reactions, combined with extended allowed reaction times or the evolution of catalytic mechanisms, would then have permitted more elaborate protein shapes, as amalgamations of earlier forms, for example.

In sum, this work extends Thusty’s (2007) elegant exploration of the evolution of the genetic code, suggesting that rate distortion considerations play a critical role across a broad spectrum of evolutionary molecular phenomena.

7 References

- Andre, I., C. Strauss, D. Kaplan, P. Bradley, and D. Baker, 2008, Emergence of symmetry in homooligomeric biological assemblies, *Proceedings of the National Academy of Sciences*, 105:16148-16152.
- Anfinsen, C., 1973, Principles that govern the folding of protein chains, *Science*, 181:223-230.
- Ash, R., 1990, *Information Theory*, Dover, New York.

- Bennett, C., 1988, Logical depth and physical complexity. In Herkin, R. (ed.), *The Universal Turing Machine: A Half-Century Survey*, Oxford University Press, pp. 227-257.
- Bos, R., 2007, Continuous representations of groupoids. arXiv:math/0612639.
- Buneci, M., 2003, *Representare de Groupoizi*, Editura Mirton, Timosoara, Romania.
- Canfield, D., M. Rosing, and C. Bjerrum, 2006, Early anaerobic metabolisms, *Philosophical Transactions of the Royal Society, B*, 351:1819-1836.
- Cover, T., and H. Thomas, 1991, *Elements of Information Theory*, Wiley, New York.
- Dembo, A., and O. Zeitouni, 1998, *Large Deviations and Applications*, 2nd edition, Springer, New York.
- Ellis, R., 1985, *Entropy, Large Deviations, and Statistical Mechanics*, Springer, New York.
- Feynman, R., 2000, *Lectures on Computation*, Westview, New York.
- Goodsell, D., and A. Olson, 2000, Structural symmetry and protein function, *Annual Reviews of Biophysics and Biomolecular Structure*, 29:105-153.
- Khinchin, A., 1957, *Mathematical Foundations of Information Theory*, Dover, New York.
- Landau, L., and E. Lifshitz, 2007, *Statistical Physics, Part I*, Elsevier, New York.
- Lei, J., and K. Huang, 2010, Protein folding: A perspective from statistical physics. arXiv:10025013v1.
- Pettini, M., 2007, *Geometry and Topology in Hamiltonian Dynamics*, Springer, New York.
- Rockafellar, R., 1970, *Convex Analysis*, Princeton University Press, Princeton, NJ.
- Thlusty, T., 2007, A model for the emergence of the genetic code as a transition in a noisy information channel, *Journal of Theoretical Biology*, 249:331-342.
- Wallace, R., and D. Wallace, 2008, Punctuated equilibrium in statistical models of generalized coevolutionary resilience: how sudden ecosystem transitions can entrain both phenotype expression and Darwinian selection, *Transactions on Computational Systems Biology IX*, LNBI 5121:23-85.
- Wallace, R., 2009, Metabolic constraints on the eukaryotic transition, *Origins of Life and Evolution of Biospheres*, 39:165-176.
- Wallace, R., 2010, Metabolic constraints on the evolution of genetic codes: Did multiple 'preaerobic' ecosystem transitions entrain richer dialects via Serial Endosymbiosis? <http://precedings.nature.com/documents/4120/version/3>.
- Weinstein, A., 1996, Groupoids: unifying internal and external symmetry, *Notices of the American Mathematical Association*, 43:744-752.
- Wolynes, P., 1996, Symmetry and the energy landscapes of biomolecules, *Proceedings of the National Academy of Sciences*, 93:14249-14255.