# A formal approach to the molecular fuzzy lock-and-key


Rodrick Wallace
Division of Epidemiology
The New York State Psychiatric Institute *


May 27, 2013


## Abstract

The fuzzy lock-and-key (FLK) powers a vast array of sophisticated logic gates at inter- and intra-cellular levels. We invoke representations of groupoid tiling wreath products analogous to the study of nonrigid molecules – or of related fuzzy symmetry extensions – to build a Morse Function that can describe spontaneous symmetry breaking phase transitions driven by information catalysis. The Function can, however, also be used to construct an Onsager-like stochastic dynamics, linked to the phase transition approach by the rich stability criteria associated with stochastic differential equations. The two methods provide complementary ways of looking at the FLK. A limit condition emerging from the stochastic dynamics gives insight into a cellular 'generalized inflammation' requiring progressively higher commitment of metabolic free energy for maintenance of basic FLK processes. These results suggest that more systematic study may illuminate pathologies associated with the failure of the FLK, a centrally-important but enigmatic biological process.



**Key Words:** cellular cognition, diffusion, glycosynapse, information theory, intrinsically disordered protein


## 1 Introduction

Since Adelman's (1994) pioneering DNA-based solution to the directed Hamiltonian path problem, a vast effort has been directed at producing molecular analogs to the usual NOT, AND, XOR, and similar logic gates, and at constructing systems using them (e.g., Stojanovic et al., 2002; Macdonald et al., 2006). Here, we will examine far more subtle naturally-occurring logic gates associated with the molecular fuzzy lock-and-key. Some background is, however, necessary.

Humberto Maturana's seminal 1970 paper, *The Biology of Cognition* introduced a perspective regarding the living state that focuses on cognition, not only high order 'neural' process, but as a phenomenon that must act at all levels of biological organization. This perspective, it will appear, provides a basis for a more comprehensive treatment of these matters.

Something of Maturana's ideas can be paraphrased as follows (Maturana, 1970; Maturana and Varela 1980, 1992):

Each internal state of a living system requires that certain interactions with the environment be satisfied in order for the system to persist. This implies that the prediction that an interaction took place once also implies it will take place again. The predictions implied in the organization of the living system are not predictions of particular events, but of classes of interactions. While every interaction is, of course, a particular interaction, every prediction is a prediction of a class of interactions that will allow the living system to retain its organization. This makes living systems inferential systems, and their domains of interactions a cognitive domain.

In consequence, living systems are cognitive systems and living as a process is a process of cognition for all organisms, with and without a nervous system. The nervous system expands the cognitive domain of the living system by making possible interactions. It does not create cognition. The nervous system, by expanding the domain of interactions of the organism, has transformed the unit of interactions and has subjected that expanded unit to the process of evolution.

Accordingly, for every living system the process of cognition consists in the creation of a field of behavior through its actual conduct in its closed domain of interactions. Consequently, although due to the historical transformations they have caused in organisms, or in their nervous systems if they have them, past interactions determine the inductive inferences that these make in the present, they do not participate in the inductive process itself.

Perhaps the first systematic information-theoretic application of these ideas was the Atlan/Cohen 'cognitive paradigm' for the immune system.

## 2 Immune cognition

Atlan and Cohen (1998) proposed an information-theoretic cognitive model of immune function and process, a paradigm incorporating cognitive pattern recognition-and-response behaviors that are certainly analogous to, but much slower than, those of the later-evolved central nervous system.

From the Atlan/Cohen perspective, the meaning of an antigen can be reduced to the type of response the antigen generates. That is, the meaning of an antigen is functionally defined by the response of the immune system. The meaning of an antigen to the system is discernible in the type of immune response produced, not merely whether or not the


*Contact: Wallace@nyspi.columbia.edu, Box 47, NYSPI, 1051 Riverside Dr., New York, NY, 10032, USA





PeerJ PrePrints | https://peerj.com/preprints/20v2/ | v2 received: 27 May 2013, published: 28 May 2013, doi: 10.7287/peerj.preprints.20v2

antigen is perceived by the receptor repertoire. Because the meaning is defined by the type of response there is indeed a response repertoire and not only a receptor repertoire.

To account for immune interpretation, Cohen (1992, 2000) has reformulated the cognitive paradigm for the immune system. The immune system can respond to a given antigen in various ways, it has 'options'. Thus the particular response observed is the outcome of internal processes of weighing and integrating information about the antigen.

In contrast to Burnet's view of the immune response as a simple reflex, it is seen to exercise cognition by the interpolation of a level of information processing between the antigen stimulus and the immune response. A cognitive immune system organizes the information borne by the antigen stimulus within a given context and creates a format suitable for internal processing; the antigen and its context are transcribed internally into the chemical language of the immune system.

The cognitive paradigm suggests a language metaphor to describe immune communication by a string of chemical signals. This metaphor is apt because the human and immune languages can be seen to manifest several similarities such as syntax and abstraction. Syntax, for example, enhances both linguistic and immune meaning.

Although individual words and even letters can have their own meanings, an unconnected subject or an unconnected predicate will tend to mean less than does the sentence generated by their connection.

The immune system creates a language by linking two ontogenetically different classes of molecules in a syntactical fashion. One class of molecules are the T and B cell receptors for antigens. These molecules are not inherited, but are somatically generated in each individual. The other class of molecules responsible for internal information processing is encoded in the individual's germline.

Meaning, the chosen type of immune response, is the outcome of the concrete connection between the antigen subject and the germline predicate signals.

The transcription of the antigens into processed peptides embedded in a context of germline ancillary signals constitutes the functional language of the immune system. Despite the logic of clonal selection, the immune system does not respond to antigens as they are, but to abstractions of antigens-in-context.

## 3 Cognition as an information source

Atlan and Cohen (1998) argue that cognition involves comparison of a perceived signal with an internal, learned or inherited picture of the world, and then, upon that comparison, choice of one response from a much larger repertoire of possible responses. Such 'choice', by reducing uncertainty, inevitably involves the transmission of information.

That is, cognitive pattern recognition-and-response proceeds by an algorithmic combination of an incoming external sensory signal with an internal ongoing activity – incorporating the internalized picture of the world – and triggering an appropriate action based on a decision that the pattern of sensory activity requires a response.

Incoming sensory input is thus mixed in an unspecified but systematic manner with internal signals to create a combined path $x = (a_0, a_1, ..., a_n, ...)$. Each $a_k$ thus represents some functional composition of the internal and the external. An application of this perspective to a standard neural network is given in Wallace (2005, p.34).

This path is fed into a highly nonlinear, but otherwise similarly unspecified, decision function, $h$, generating an output $h(x)$ that is an element of one of two disjoint sets $B_0$ and $B_1$ of possible system responses. Let

$$B_0 \equiv \{b_0, ..., b_k\},$$

$$B_1 \equiv \{b_{k+1}, ..., b_m\}.$$

Assume a graded response, supposing that if $h(x) \in B_0$, the pattern is not recognized, and if $h(x) \in B_1$, the pattern is recognized, and some action $b_j, k+1 \leq j \leq m$ takes place.

Interest focuses on those paths $x$ triggering pattern recognition-and-response: given a fixed initial state $a_0$, examine all possible subsequent paths $x$ beginning with $a_0$ and leading to the event $h(x) \in B_1$. Thus $h(a_0, ..., a_j) \in B_0$ for all $0 \leq j < m$, but $h(a_0, ..., a_m) \in B_1$.

For each positive integer $n$, let $N(n)$ be the number of high probability paths of length $n$ that begin with some particular $a_0$ and lead to the condition $h(x) \in B_1$. Call such paths 'meaningful', assuming that $N(n)$ will be considerably less than the number of all possible paths of length $n$ leading from $a_0$ to the condition $h(x) \in B_1$.

Note that identification of the 'alphabet' of the states $a_j, B_k$ may depend on the proper system 'coarse graining' in the sense of symbolic dynamics (Beck and Schlogl, 1993).

Combining algorithm, the form of the function $h$, and the details of grammar and syntax, are all unspecified in this model. The assumption permitting inference on necessary conditions constrained by the asymptotic limit theorems of information theory is that the finite limit

$$H \equiv \lim_{n \to \infty} \frac{\log[N(n)]}{n}$$

both exists and is independent of the path $x$.

Call such a pattern recognition-and-response cognitive process *ergodic*. Not all cognitive processes are likely to be ergodic, implying that $H$, if it indeed exists at all, is path dependent, although extension to nearly ergodic processes, in a certain sense, seems possible (e.g., Wallace, 2005, pp. 31-32).

Invoking the spirit of the Shannon-McMillan Theorem, it is possible to define an adiabatically, piecewise stationary, ergodic information source $\mathbf{X}$ associated with stochastic variates $X_j$ having joint and conditional probabilities $P(a_0, ..., a_n)$ and $P(a_n|a_0, ..., a_{n-1})$ such that appropriate joint and conditional Shannon uncertainties satisfy the relations:

$$H[X] = \lim_{n \to \infty} \frac{\log[N(n)]}{n} =$$

PeerJ PrePrints | https://peerj.com/preprints/20v2/ | v2 received: 27 May 2013, published: 28 May 2013, doi: 10.7287/peerj.preprints.20v2

$$\lim_{n\to\infty} H(X_n|X_0,...,X_{n-1}) = \lim_{n\to\infty} \frac{H(X_0, X_1, ..., X_n)}{n+1} . \quad (1)$$

Shannon uncertainties all have the form $-\sum_k P_k \log(P_k)$, with $\sum_k P_k = 1$ (Cover and Thomas, 2006).

The information source $H[X]$ is defined as *dual* to the underlying ergodic cognitive process.

'Adiabatic' means that, when the information source is parameterized according to some appropriate scheme, within continuous 'pieces', changes in parameter values take place slowly enough so that the information source remains as close to stationary and ergodic as needed to make the fundamental limit theorems work. 'Stationary' means that probabilities do not change in time, and 'ergodic' (roughly) that cross-sectional means converge to long-time averages. Between 'pieces' one invokes various kinds of phase change formalism, for example renormalization theory in cases where a mean field approximation holds (Wallace, 2005), or variants of random network theory where a mean number approximation is applied. This will be outlined below.

An equivalence class algebra can be constructed by choosing different origin points, $a_0$, and defining the equivalence of two states, $a_m, a_n$, by the existence of high probability meaningful paths connecting them to the same origin point. Disjoint partition by equivalence class, analogous to orbit equivalence classes for dynamical systems, defines the vertices of a network of cognitive dual languages. Each vertex then represents a different information source dual to a cognitive process. This is not a representation of a cognitive network as such. It is, rather, an abstract set of 'languages' dual to the set of interacting cognitive processes.

This structure generates a groupoid, an extension of the idea of the symmetry group, as summarized by Brown (1987) and Weinstein (1996). See the Mathematical Appendix for an introduction to groupoids, which will appear later in the argument as generalizations of simple molecular symmetries applied to tilings that are only locally symmetric.

# 4 Information catalysis

Information, although a form of free energy (Feynman, 2000), per se does not itself carry very much ability to do work, but the physical mechanisms that instantiate signals do, and this fact, in concert with the asymptotic limit theorems of information theory, permits an important general argument.

Suppose there are two interacting information sources $X, Y$, emitting sequences of signals $x = [x_1, x_2, ...]$ and $y = [y_1, y_2, ...]$ at times $i = 1, 2, ....$ A joint sequence of signals $xy \equiv [(x_1, y_1), (x_2, y_2), ...]$ can then be defined, and, where the individual sequences $x$ and $y$ are correlated, it is possible to define a joint source information source uncertainty $H_{X,Y}$ for which a version of the information theory chain rule applies (Cover and Thomas, 2006):

$$H_{X,Y} < H_X + H_Y . \quad (2)$$

The average production of information, $\hat{H}$, from a process having an available metabolic free energy rate $M$, can be expected to follow a relation having the standard Gibbs form

$$\hat{H} = \frac{\int H \exp[-H/\kappa M] dH}{\int \exp[-H/\kappa M] dH} \approx \kappa M , \quad (3)$$

where $\kappa$ is quite small, so the integral converges.

Then, from the chain rule,

$$\begin{aligned} \hat{H}_{X,Y} &< \hat{H}_X + \hat{H}_Y , \\ M_{X,Y} &< M_X + M_Y . \end{aligned} \quad (4)$$

If $X$ is the system of interest, then, at the expense of maintaining the regulatory information source $Y$, it is possible to canalize the reaction paths of $X$: $M_{X,Y}$ becomes a valley in the larger energy structure created by imposing $Y$ and $X$ together.

Interpreting the metabolic free energy rate $M_Y$ driving the information source $Y$ as a temperature analog, we are now in a position to examine more complicated biological logic gates than are currently described in the computing literature.

# 5 The fuzzy lock-and-key

The fuzzy lock-and-key dominates many mechanisms that transmit information at inter- and intra-cellular levels. Indeed, 30% of all proteins are 'intrinsically disordered' (IDP), and, by some measures, perhaps 50% of all proteins have significant regions that are intrinsically disordered. Such structure – or rather, its lack – allows operation of the extraordinarily flexible logic gates necessary for many of the cognitive processes that are the foundation of the living state (e.g., Maturana, 1970). Figure 1, adapted from Tompa et al. (2005), provides and example in which the same IDP can either activate or inhibit a chemical logic gate, depending on an 'information catalysis' in which an incoming signal splits isoenergetic groupoid tiling symmetry states via an analog to spontaneous symmetry breaking, making one or the other the lower energy conformation (e.g., Wallace, 2011a, 2012). Far more sophisticated logic gates can easily be constructed quite simply using similar mechanisms.

Figure 2 shows another example, a frond of the highly flexible 'glycan kelp bed' that coats the cell surface, and, via binding with lectins, triggers even more complicated logical processes. While proteins are constructed from 20 basic amino acids, the glycan kelp bed is formed from as many as 7,000 glycan determinants, and represents a vastly more complex system for information transmission (Cummings, 2009; Gupta et al., 2010).

Figure 3, from Dam et al. (2007), illustrates a 'bind-and-slide' mechanism by which increasing concentration of a lectin species can induce a phase transition topological change. Initially, the lectin diffuses along and off the glycan kelp frond, until a sufficient number of sites are occupied. Then the lectin-coated fronds cross bind until the reaction saturates, triggering the gate.



PeerJ PrePrints | https://peerj.com/preprints/20v2/ | v2 received: 27 May 2013, published: 28 May 2013, doi: 10.7287/peerj.preprints.20v2

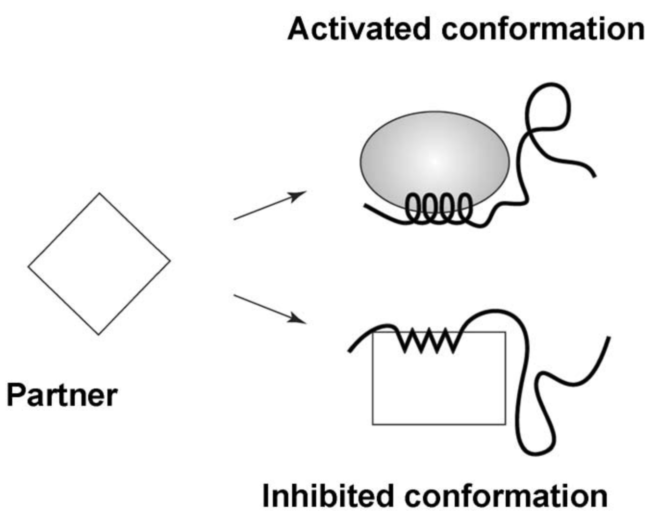


Figure 1: From Tompa et al., 2005. The partner can bind in two ways to the IDP. The top form is activated, and the bottom inhibited. The triggering between the states is done by an 'information catalysis' in which an incoming signal shifts the lowest energy state between the two otherwise thermodynamically competitive – isoenergetic – topological forms via a kind of spontaneous symmetry breaking acting on tiling groupoids.

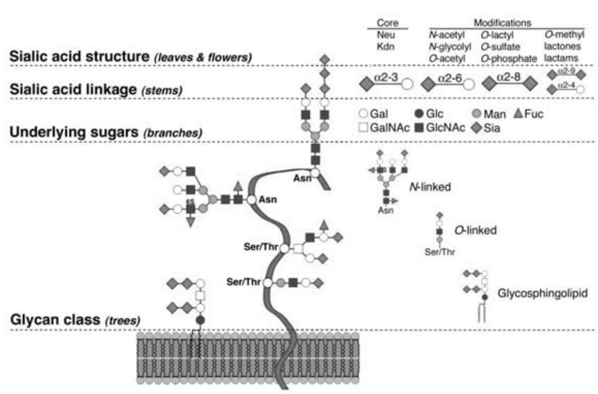


Figure 2: From Cohen and Varki, (2010). Levels of sialome complexity, from core and core modifications to the shifting, bending, twisting, glycan 'kelp fronds' that coat most cell surfaces and, via lectin interaction, constitute sophisticated logic gates involved in explosively vast information transfers: in comparison with the 20 amino acids making up all proteins, some 7,000 glycan determinants are needed to constitute the flexible kelp fronds, side branches and all (Cummings, 2009).

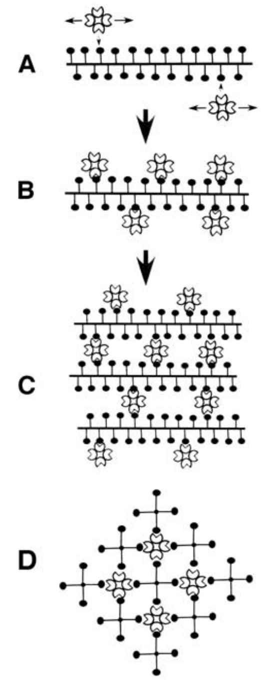


Figure 3: From Dam et al. (2007). Lectin diffuses along and off the flexible glycan frond, until a sufficient number of sites are occupied. Then the coated glycan fronds begin to cross bind, the reaction is complete, and the logic gate is activated. The last figure shows an end view.

Wallace (2011a, 2012) applies nonrigid molecule symmetries to IDP, and Wallace and Wallace (2013, chapter 8) extend the analysis to the glycan/lectin interface. Here we will generalize the argument across chemical species, and examine what may be an important stability criterion that appears to underlie all possible such mechanisms.

We begin with a brief recapitulation of the basic formalism.

# 6 Symmetries of the FLK

One basis for the approach is the classic observation by Longuet-Higgins (1963) that the symmetry group of a nonrigid molecule is the set of (i) all feasible permutations of the positions and spins of identical nuclei and (ii) all feasible permutation-inversions, which simultaneously invert the coordinates of all particles in the center of mass.

It may then, for some forms of the FLK, be possible to extend nonrigid molecule group theory using wreath, semidirect, or other products over a set of finite and/or compact groups (e.g., Balasubramanian, 1980, 2004), or their groupoid generalizations, as now common in stereochemistry (Wallace, 2011b and cited references). Groupoids are local structures that characterize the partial symmetries of finite tilings, quasicrystals, and the like, and provide a highly natural means of extending local symmetries (Brown, 1987; Weinstein, 1996). The simplest groupoid can be envisioned as a disjoint union of groups, so that the group element product is only locally defined. In addition, equivalence classes define groupoids, so that the concept generalizes both structures.

The groups or groupoids of interest are taken as parameterized by an index of 'topological complexity', in a large sense, a temperature-analog $L$. In general, the number of group/groupoid elements can be expected to grow exponentially with $L$, typically as $\sum \Pi_j |G_j||A_j|^L$, where $|G_k|$ and $|A_k|$ are the size, in an appropriate sense, of symmetry groups $G_k$ and $A_k$. See the Balasubramanian references for details.

Kahraman (2009) argues that the observed 'sloppiness' of large lock/small key molecular reaction dynamics suggests that binding site symmetry may be greater than binding ligand symmetries. Thus binding ligands may be expected to involve dual, mirror subgroups/groupoids of the anchored nonrigid group/groupoid symmetries of the binding site. Thus the argument becomes:

> Increasing $L, |G|, |A|$ $\rightarrow$ more flexibility $\rightarrow$ greatly enlarged binding site nonrigid symmetry group/groupoid $\rightarrow$ more subgroups/subtilings of possible binding sites for ligand attachment.

This can be addressed by supposing that the duality between a subgroup or subgroupoid of the fuzzy lock and of the fuzzy key can be expressed as

$$\mathcal{B}_\alpha = C_\beta \mathcal{D}_\gamma \tag{5}$$

where $\mathcal{B}_\alpha$ is a subgroup/groupoid (or set of them) of the appropriate nonrigid symmetry group or groupoid, $\mathcal{D}_\gamma$ a similar structure of the set of binding ligands, and $C_\beta$ is an appropriate inversion operation or set of them that represents static or dynamic matching between them. The fuzziness, Wallace and Wallace (2013) argue, can even extend to sequence replacement as well as geometric variations.

An outcome of this approach is that FLK matching symmetries, and their associated dynamics, can be highly punctuated in the parameter $L$ that broadly indexes topological complexity.

A nonrigid molecule analog based on wreath products of tiling groupoids is not the only possible attack on the FLK problem. Paul Mezey and colleagues have introduced another extension of simple molecular symmetries using a fuzzy set approach (e.g., Mezey, 1997). In that methodology, the sharply defined families of nuclear arrangements with specified point symmetry are replaced by fuzzy sets – so-called 'syntopy sets' – of arrangements having only some degree of symmetry of the original perfect point symmetries. The method provides the syntopy sets with a group theoretic characterization, and the syntopy groups retain some aspects of the underlying point groups, gaining, however, a continuous parameterization. Mezey further generalizes these ideas to what he calls fuzzy symmorphy groups.

In essence, the 'fuzzification' of algebraic structures and relations is based on an extension of the characteristic function, mapping an arbitrary set into the set of integers $0, 1$, so that $f : G \rightarrow 0, 1$. Then, if $x \in G, f(x) = 1$, otherwise $f(x) = 0$. Generalization involves letting $f$ map onto the real interval $[0, 1]$. Rosenfeld (1971) applied the method to groups and groupoids, and application to group/groupoid representations seems direct, albeit modified by some of the complexities associated with groupoid wreath products and other matters (Houghton, 1975; Bos, 2007).

To the extent that representations of these objects are possible, the Morse Function techniques that follow should carry through.

# 7 Phase transitions and reaction dynamics

Again, interpret the regulatory free energy intensity $M_Y$ associated with an information catalyst having an information source $Y$ as a pseudo-temperature index $\mathcal{T}$. For large $\mathcal{T}$, it becomes possible to apply a statistical mechanics analog, and to use Landau's spontaneous symmetry breaking/lifting approach via a Morse Theory argument (Wallace, 2012; Pettini, 2007). See the Mathematical Appendix for a summary of standard material on Morse Theory. Typically, very many Morse functions are possible under a given circumstance, and it is possible to construct what is perhaps the simplest using representations of the appropriate generalized groupoids and/or groups. Although representations of groupoids are, broadly, similar to those of groups, there are necessary modifications (Bos, 2007).

Taking an appropriate group (or groupoid) representation in a particular matrix or function algebra, now construct a 'pseudo probability' $\mathcal{P}$ for nonrigid group element $\omega$ as

$$\mathcal{P}[\omega] = \frac{\exp[-|\chi_\omega|/\kappa\mathcal{T}]}{\sum_\nu \exp[-|\chi_\nu|/\kappa\mathcal{T}]} . \tag{6}$$

$\chi_\phi$ is the character of the group element $\phi$ in that representation, i.e., the trace of the matrix or function assigned to $\phi$, and $|...|$ is the norm of the character, a real number. For systems that include compact groups, the sum may be a generalized integral.

The central idea is that $F$ in the construct

$$\exp[-F/\kappa\mathcal{T}] = \sum_\nu \exp[-|\chi_\nu|/\kappa\mathcal{T}] \tag{7}$$

is a Morse Function in the signaling temperature-analog $\mathcal{T}$ to which Landau's spontaneous symmetry breaking arguments apply (Wallace, 2012; Pettini, 2007; Landau and Lifshitz, 2007). This leads to the expectation of empirically observable highly punctuated structure and reaction dynamics in the index $\mathcal{T}$ that are the analog to phase transitions in 'simple' physical systems.

Recall Landau's central insight: for many physical phenomena, raising the temperature makes accessible higher energy states of the system Hamiltonian, the quantum mechanical energy operator, and the inherent symmetry changes are necessarily be punctuated. Here the focus is directly on a Morse Function constructed from a representation of underlying nonrigid groupoid wreath product tiling symmetries.

However, topological matters – the shape of a system has – long been known to profoundly affect phase transition behavior (e.g., Privman and Fisher, 1983). Thus, a distinctly different approach is also possible to FLK reaction mechanism. The basic assumption is that the group or groupoid tiling symmetries of the fuzzy lock must be matched by an appropriate set of keys in a dynamic manner. Thus the statistical mechanics of fuzzy interaction symmetries becomes central to reaction trajectories, treated here according to an Onsager-like nonequilibrium thermodynamics formulation.

Define, then, a 'symmetry entropy' based on the Morse Function $F$ of equation (7) over a set of underlying structural or other parameters $\mathbf{Q} = [Q_1, ..., Q_n]$ as the Legendre transform

$$S = F(\mathbf{Q}) - \sum_i Q_i \partial F(\mathbf{Q})/\partial Q_i \,. \qquad (8)$$

The time behavior of such a system will be driven, at least in first approximation, by standard Onsager-like nonequilibrium thermodynamics relations (de Groot and Mazur, 1984):

$$dQ_i/dt = \sum_j \mathcal{K}_{i,j} \partial S/\partial Q_j \,, \qquad (9)$$

where the $\mathcal{K}_{i,j}$ are appropriate empirical parameters and $t$ is the time. The system may, or may not, have local time reversibility. If not, then $\mathcal{K}_{i,j} \neq \mathcal{K}_{j,i}$.

Since, however, this is essentially a 'fuzzy' system, a more fitting approach is through a set of stochastic differential equations having the form:

$$dQ_t^i = \mathcal{K}_i(t, \mathbf{Q})dt + \sum_j \sigma_{i,j}(t, \mathbf{Q})dB^j \,, \qquad (10)$$

where the $\mathcal{K}_i$ and $\sigma_{i,j}$ are appropriate functions.

Different kinds of 'noise' $dB^j$ will have particular forms of quadratic variation affecting dynamics.

Setting the expectation of this equation to zero and solving for stationary points gives attractor states, since noise precludes unstable equilibria, although the solution may, in fact, be a highly dynamic strange attractor set.

But setting the expectation of equation (10) to zero also generates an index theorem (Hazewinkel, 2002) in the sense of Atiyah and Singer (1963) that relates analytic results – the solutions of the equations – to an underlying set of topological structures representing the eigenmodes of a complicated 'nonrigid molecule' geometric operator whose group/groupoid spectrum represents the symmetries of the possible FLK reactions that must take place for information to be transmitted, i.e., for the chemical logic gate to be triggered.

A one-dimensional model, however, raises significant questions regarding the stability of the dynamics of the fuzzy lock-and-key in the presence of noise. This will be a specific example of a well-known general phenomenon: systems described by stochastic differential equations can be stable in the expectation of the first moment, the mean, but may be unstable in the expectation of some higher moment, triggering a catastrophe (Khasminskii, 2010). That catastrophe appears analogous to spontaneous symmetry breaking in the metabolic free energy rate index $\mathcal{T}$.

# 8 An example

The motivation for this approach is as follows. Regulation can be viewed in terms of the average distortion between signals sent by the regulating agent and the observed impact on the regulated system. The Rate Distortion Function (RDF), $R(D)$, measures the minimum signal channel capacity – a free energy index – needed to keep the average distortion less than or equal to some value $D$, using a particular distortion measure (Cover and Thomas, 2006). For a Gaussian channel under the squared measure, $R(D) = 1/2 \log[\sigma^2/D]$, where $\sigma^2$ is the variance of the inherent channel noise. Define an 'RDF entropy' as

$$S_R = R(D) - DdR/dD = 1/2 \log[\sigma^2/D] + 1/2 \,. \qquad (11)$$

The simplest nonequilibrium Onsager equation is just

$$dD/dt = -\mu dS_R/dD = \mu/2D \,, \qquad (12)$$

where $t$ is the time and $\mu$ the diffusion coefficient. By inspection,

$$D(t) = \sqrt{\mu t} \,. \qquad (13)$$

This is the classic solution to the diffusion equation, a correspondence reduction to a well-known result that can serve as a basis for arguing upward in complexity.

Regulation does not involve diffusive drift, but rather consumes massive amounts of free energy at high rates to ensure that target systems operate within characteristic limits. Let $G(\mathcal{T})$ represent a monotonic increasing function of the rate of free energy consumption $\mathcal{T}$, then a plausible form of equation (10), in the presence of an added regulatory system noise indexed by $\beta/2$, is

$$dQ_t = [f(Q_t) - G(\mathcal{T})]dt + \frac{\beta}{2} Q_t dW_t \,, \qquad (14)$$

where $dW_t$ is standard white noise, $G(\mathcal{T})$ is as above, and the last term represents a volatility effect.

This has the simple equilibrium expectation

$$Q_{equlib} = f^{-1}(G(\mathcal{T})) \,. \qquad (15)$$

However, the presence of the noise term can introduce serious complications. Suppose, following the example of equation (12), $f = \alpha/Q$. Then determining the variance of $Q$ involves using the Ito chain rule on the variate $Y = Q^2$. This leads to the stochastic differential equation

$$dY_t = [2\sqrt{Y_t}(\frac{\alpha}{\sqrt{Y_t}} - G(\mathcal{T})) + \frac{\beta^2}{4} Y_t]dt + \beta Y_t dW_t \,, \qquad (16)$$

where $(\beta^2/4)Y_t$ in the time term is the Ito correction.

Taking the expectation at equilibrium gives a condition for a real solution for the variance of $Q$ involving the discriminant of a quadratic equation:

$$G(\mathcal{T}) > \beta\sqrt{\alpha/2} \,. \qquad (17)$$

If this condition is not satisfied, then there can be no real expectation in the second moment of $Q$.



PeerJ PrePrints | https://peerj.com/preprints/20v2/ | v2 received: 27 May 2013, published: 28 May 2013, doi: 10.7287/peerj.preprints.20v2

Note that if $f \propto 1/A_n(Q)$, where $A_n$ is a polynomial of degree $n$, then the equivalent of equation (17) will involve the discriminant of a polynomial of degree $n+1$.

$G(\mathcal{T})$ determines the metabolic free energy needed to activate FLK dynamics, in this model. Solving for $\mathcal{T}$ gives

$$\mathcal{T} > G^{-1}(\beta\sqrt{\alpha/2})\,. \tag{18}$$

Taking a Landau spontaneous symmetry breaking perspective, $\mathcal{T}$ in equation (18) represents the minimum rate of free energy expenditure needed to maintain a high state of symmetry in the FLK system. Lowering $\mathcal{T}$ under that limit triggers a phase transition to a simpler, disjointed, nonfunctional – or at least differently functional – structure, potentially a catastrophe, but at the very least, a different reaction regime.

Depending on the form of $G^{-1}$, small increase in $\beta$ may cause significant increase in the free energy needed to properly control FLK dynamics, according to the model. Such an event could represent a kind of generalized inflammation, a persistent overdrive, that could cause long-term physiological damage, as does chronic activation of the immune system (e.g., Wallace and Wallace, 2010, 2013, and references therein).

More generally, however, the rich stability criteria associated with systems described by equation (10) may provide tools for understanding a broad class of symmetry changes across the dynamics of the FLK, not just those of catastrophic failure. This could give a method for exploring the spectrum determined by the underlying Atiyah/Singer index theorem associated with equation (10).

# 9 Discussion and conclusions

The fuzzy lock-and-key drives a vast array of elaborate logic gates at inter- and intra-cellular levels of biological structure. Indeed, the glycan kelp bed that coats the cell surface provides one of the most information-rich of biological environments (Gupta et al., 2010), one that Cohen and Varki (2010) characterize in terms of a 'glycosynapse' that apparently rivals the neural synapse in sophistication. While there may be some $10^{11}$ active neurons in humans, virtually all living cells within an organism may have numerous glycosynapses engaging in complicated information switching. Within cells there are even more FLK logic gates using IDP, or using regions of structured proteins that are intrinsically disordered. Thus the numbers of FLK logic gates within an organism are literally astronomical, far more numerous than neural synapses. This might well be called the Maturana world of the organism.

Here, we have used representations of groupoid tiling wreath products, or other possible symmetry descriptions associated with the FLK, to construct a Morse Function that can describe both spontaneous symmetry breaking phase transitions driven by information catalysis, and can be used to construct an Onsager-like stochastic dynamics. The two approaches appear linked by the rich instability structure possible to stochastic differential equations.

The limit condition of equation (18) may, in addition, give insight into a cellular 'generalized inflammation' requiring higher and higher commitment of metabolic free energy for maintenance of basic FLK processes, leading to pathologies analogous to those resulting from overactive immune or HPA axis systems (e.g., Wallace and Wallace, 2010). A more complete study may provide a deeper understanding of the broad array of serious dysfunctions that must inevitably be associated with failures of the FLK's, since these are among the most basic phenomena of the living state.

# 10 Mathematical Appendix

## 10.1 Morse Theory

Morse Theory examines relations between analytic behavior of a function – the location and character of its critical points – and the underlying topology of the manifold on which the function is defined. Here we follow Pettini (2007).

The underlying idea of Morse Theory is to examine an $n$-dimensional manifold $M$ as decomposed into level sets of some function $f : M \to \mathbf{R}$ where $\mathbf{R}$ is the set of real numbers. The $a$-level set of $f$ is defined as

$$f^{-1}(a) = \{x \in M : f(x) = a\},$$

the set of all points in $M$ with $f(x) = a$. If $M$ is compact, then the whole manifold can be decomposed into such slices in a canonical fashion between two limits, defined by the minimum and maximum of $f$ on $M$. Let the part of $M$ below $a$ be defined as

$$M_a = f^{-1}(-\infty, a] = \{x \in M : f(x) \le a\}.$$

These sets describe the whole manifold as $a$ varies between the minimum and maximum of $f$.

Morse functions are defined as a particular set of smooth functions $f : M \to \mathbf{R}$ as follows. Suppose a function $f$ has a critical point $x_c$, so that the derivative $df(x_c) = 0$, with critical value $f(x_c)$. Then, $f$ is a Morse function if its critical points are nondegenerate in the sense that the Hessian matrix of second derivatives at $x_c$, whose elements, in terms of local coordinates are

$$\mathcal{H}_{i,j} = \partial^2 f/\partial x^i \partial x^j,$$

has rank $n$, which means that it has only nonzero eigenvalues, so that there are no lines or surfaces of critical points and, ultimately, critical points are isolated.

The index of the critical point is the number of negative eigenvalues of $\mathcal{H}$ at $x_c$.

A level set $f^{-1}(a)$ of $f$ is called a critical level if $a$ is a critical value of $f$, that is, if there is at least one critical point $x_c \in f^{-1}(a)$.

Again following Pettini (2007), the essential results of Morse Theory are:

1. If an interval $[a, b]$ contains no critical values of $f$, then the topology of $f^{-1}[a, v]$ does not change for any $v \in (a, b]$. Importantly, the result is valid even if $f$ is not a Morse function, but only a smooth function.


PeerJ PrePrints | https://peerj.com/preprints/20v2/ | v2 received: 27 May 2013, published: 28 May 2013, doi: 10.7287/peerj.preprints.20v2

2. If the interval $[a, b]$ contains critical values, the topology of $f^{-1}[a, v]$ changes in a manner determined by the properties of the matrix $H$ at the critical points.

3. If $f : M \rightarrow \mathbf{R}$ is a Morse function, the set of all the critical points of $f$ is a discrete subset of $M$, i.e., critical points are isolated. This is Sard's Theorem.

4. If $f : M \rightarrow \mathbf{R}$ is a Morse function, with $M$ compact, then on a finite interval $[a, b] \subset \mathbf{R}$, there is only a finite number of critical points $p$ of $f$ such that $f(p) \in [a, b]$. The set of critical values of $f$ is a discrete set of $\mathbf{R}$.

5. For any differentiable manifold $M$, the set of Morse functions on $M$ is an open dense set in the set of real functions of $M$ of differentiability class $r$ for $0 \leq r \leq \infty$.

6. Some topological invariants of $M$, that is, quantities that are the same for all the manifolds that have the same topology as $M$, can be estimated and sometimes computed exactly once all the critical points of $f$ are known: let the Morse numbers $\mu_i (i = 0, ..., m)$ of a function $f$ on $M$ be the number of critical points of $f$ of index $i$, (the number of negative eigenvalues of $H$). The Euler characteristic of the complicated manifold $M$ can be expressed as the alternating sum of the Morse numbers of any Morse function on $M$,

$$\chi = \sum_{i=1}^{m} (-1)^i \mu_i.$$

The Euler characteristic reduces, in the case of a simple polyhedron, to

$$\chi = V - E + F$$

where $V, E$, and $F$ are the numbers of vertices, edges, and faces in the polyhedron.

7. Another important theorem states that, if the interval $[a, b]$ contains a critical value of $f$ with a single critical point $x_c$, then the topology of the set $M_b$ defined above differs from that of $M_a$ in a way which is determined by the index, $i$, of the critical point. Then $M_b$ is homeomorphic to the manifold obtained from attaching to $M_a$ an $i$-handle, i.e., the direct product of an $i$-disk and an $(m - i)$-disk.

Pettini (2007) and Matsumoto (2002) contain mathematical details and further references.

## 10.2 Groupoids

Following Weinstein (1996), a groupoid, $G$, is defined by a base set $A$ upon which some mapping – a morphism – can be defined. Note that not all possible pairs of states $(a_j, a_k)$ in the base set $A$ can be connected by such a morphism. Those that can define the groupoid element, a morphism $g = (a_j, a_k)$ having the natural inverse $g^{-1} = (a_k, a_j)$. Given such a pairing, it is possible to define 'natural' end-point maps $\alpha(g) = a_j, \beta(g) = a_k$ from the set of morphisms $G$ into $A$, and a formally associative product in the groupoid $g_1 g_2$ provided $\alpha(g_1 g_2) = \alpha(g_1), \beta(g_1 g_2) = \beta(g_2)$, and $\beta(g_1) = \alpha(g_2)$. Then, the product is defined, and associative, $(g_1 g_2) g_3 = g_1 (g_2 g_3)$.

In addition, there are natural left and right identity elements $\lambda_g, \rho_g$ such that $\lambda_g g = g = g \rho_g$ (Weinstein, 1996).

An orbit of the groupoid $G$ over $A$ is an equivalence class for the relation $a_j \sim G a_k$ if and only if there is a groupoid element $g$ with $\alpha(g) = a_j$ and $\beta(g) = a_k$. Following Cannas da Silva and Weinstein (1999), we note that a groupoid is called transitive if it has just one orbit. The transitive groupoids are the building blocks of groupoids in that there is a natural decomposition of the base space of a general groupoid into orbits. Over each orbit there is a transitive groupoid, and the disjoint union of these transitive groupoids is the original groupoid. Conversely, the disjoint union of groupoids is itself a groupoid.

The isotropy group of $a \in X$ consists of those $g$ in $G$ with $\alpha(g) = a = \beta(g)$. These groups prove fundamental to classifying groupoids.

If $G$ is any groupoid over $A$, the map $(\alpha, \beta) : G \rightarrow A \times A$ is a morphism from $G$ to the pair groupoid of $A$. The image of $(\alpha, \beta)$ is the orbit equivalence relation $\sim G$, and the functional kernel is the union of the isotropy groups. If $f : X \rightarrow Y$ is a function, then the kernel of $f$, $ker(f) = [(x_1, x_2) \in X \times X : f(x_1) = f(x_2)]$ defines an equivalence relation.

Groupoids may have additional structure. As Weinstein (1996) explains, a groupoid $G$ is a topological groupoid over a base space $X$ if $G$ and $X$ are topological spaces and $\alpha, \beta$ and multiplication are continuous maps. A criticism sometimes applied to groupoid theory is that their classification up to isomorphism is nothing other than the classification of equivalence relations via the orbit equivalence relation and groups via the isotropy groups. The imposition of a compatible topological structure produces a nontrivial interaction between the two structures. Above, we have introduced a metric structure on manifolds of related information sources, producing such interaction.

In essence, a groupoid is a category in which all morphisms have an inverse, here defined in terms of connection to a base point by a meaningful path of an information source dual to a cognitive process.

As Weinstein (1996) points out, the morphism $(\alpha, \beta)$ suggests another way of looking at groupoids. A groupoid over $A$ identifies not only which elements of $A$ are equivalent to one another (isomorphic), but it also parameterizes the different ways (isomorphisms) in which two elements can be equivalent, i.e., in our context, all possible information sources dual to some cognitive process. Given the information theoretic characterization of cognition presented above, this produces a full modular cognitive network in a highly natural manner.

Brown (1987) describes the fundamental structure as follows:

> A groupoid should be thought of as a group with many objects, or with many identities... A groupoid with one object is essentially just a group. So the notion of groupoid is an extension of that of groups. It gives an additional convenience, flexibility and range of applications...
>
> EXAMPLE 1. A disjoint union [of groups] $G = \cup_\lambda G_\lambda, \lambda \in \Lambda$, is a groupoid: the product $ab$ is defined if and only if $a, b$ belong to the same $G_\lambda$, and $ab$ is

PeerJ PrePrints | https://peerj.com/preprints/20v2/ | v2 received: 27 May 2013, published: 28 May 2013, doi: 10.7287/peerj.preprints.20v2

then just the product in the group $G_\lambda$. There is an identity $1_\lambda$ for each $\lambda \in \Lambda$. The maps $\alpha, \beta$ coincide and map $G_\lambda$ to $\lambda$, $\lambda \in \Lambda$.

EXAMPLE 2. An equivalence relation $R$ on [a set] $X$ becomes a groupoid with $\alpha, \beta : R \to X$ the two projections, and product $(x, y)(y, z) = (x, z)$ whenever $(x, y), (y, z) \in R$. There is an identity, namely $(x, x)$, for each $x \in X$...

Weinstein (1996) makes the fundamental point that almost every interesting equivalence relation on a space $B$ arises in a natural way as the orbit equivalence relation of some groupoid $G$ over $B$. Instead of dealing directly with the orbit space $B/G$ as an object in the category $S_{map}$ of sets and mappings, one should consider instead the groupoid $G$ itself as an object in the category $G_{htp}$ of groupoids and homotopy classes of morphisms.

The groupoid approach has become quite popular in the study of networks of coupled dynamical systems which can be defined by differential equation models, (e.g., Golubitsky and Stewart 2006).

# 11 References


Adelman, L., 1994, Molecular computation of solutions to combinatorial problems, Science, 266:1021-1024.

Atiyah, M., I. Singer, 1963, The index of elliptical operators on compact manifolds, Bulletin of the American Mathematical Society, 69:322-433.

Atlan, H., I. Cohen, 1998, Immune information, self-organization, and meaning, International Immunology, 10:711-717.

Balasubramanian, K., 1980, The symmetry groups of non-rigid molecules as generalized wreath products and their representations, Journal of Chemical Physics, 72:665-677.

Balasubramanian, K., 2004, Relativistic double group spinor representations of nonrigid molecules, Journal of Chemical Physics, 120:5524-5535.

Beck, C., F. Schlogl, 1993, Thermodynamics of Chaotic Systems, Cambridge University Press, New York.

Bos, R., 2007, Continuous representations of groupoids, arXiv:math/0612639v3.

Brown, R., 1987, From groups to groupoids: a brief survey, Bulletin of the London Mathematical Society, 19:113-1134.

Cannas da Silva, A., A. Weinstein, 1999, Geometrtic Models for Noncommutative Algebras, American Mathematical Society, Providence, RI.

Cohen, I., 1992, The cognitive principle challenges clonal selection, Immunology Today, 13:441-444.

Cohen, I., 2000, Tending Adam's Garden: Evolving the Cognitive Immune Self, Academic Press, New York.

Cohen, M., A. Varki, 2010, The sialome - far more than the sum of its parts, OMICS, 14:455-464.

Cover, T., J. Thomas, 2006, Elements of Information Theory, Second Edition, Wiley, New York.

Cummings, R., 2009, The repertoire of glycan determinants in the human glycome, Molecular BioSystems, 5:1087-1104.

Dam, T. T. Gerken, B. Cavada, K. Nascimento, T. Moura, C.F. Brewer, 2007, Binding studies of the $\alpha$-GalNAc-specific lectins to the $\alpha$-GalNAc(Tn-antigen) form porcine submaxilary mucin and its smaller fragments, Journal of Biological Chemistry, 38:28256-28263.

de Groot, S., P. Mazur, 1984, Non-Equilibrium Thermodynamics, Dover, New York.

Feynman, R., 2000, Lectures on Computation, Westview Press, New York.

Golubitsky, M., I. Stewart, 2006, Nonlinear dynamics and networks: the groupoid formalism, Bulletin of the American Mathematical Society, 43:305-364.

Gupta, G., A. Surolia, S. Sampath Kumar, 2010, Lectin microarrays for glycomic analysis, OMICS, 14:419-436.

Hazewinkel, M., 2002, Encyclopedia of Mathematics, 'Index Formulas', Springer, New York.

Houghton, C., 1975, Wreath products of groupoids, Journal of the London Mathematical Society, 10:179-188.

Kahraman, A., 2009, The geometry and physicochemistry of protein binding sites and ligands and their detection in electron density maps, PhD Dissertation, Cambridge University.

Khasminskii, R., 2012, Stochastic Stability of Differential Equations, Second Edition, Springer, New York.

Landau, L., E. Lifshitz, 2007, Statistical Physics, Part I, Elsevier, New York.

Longuett-Higgins, H., 1963, The symmetry groups of non-rigid molecules, Molecular Physics, 6:445-460.

Macdonald, J., L. Yang, M. Sutovic, H. Lederman, K. Pendri, W. Lu, B. Andrews, D. Stefanovic, M. Stojanovic, 2006, Medium scale integration of molecular logic gates in an automaton, Nano Letters, 6:2598-2603.

Matsumoto, Y., 2002, An Introduction to Morse Theory, American Mathematical Society, Providence.

Maturana, H., 1970, The biology of cognition, BCL Research Report 9.0, University of Illinois.

Maturana, H., F. Varela, 1980, Autopoiesis and Cognition, Reidel Publishing Company, Dortrecht.

Maturana, H., F. Varela, 1992, The Tree of Knowledge, Shambhala Publications, Boston.

Mezey, P., 1997, Fuzzy measures of molecular shape and size, Chapter 5 in Fuzzy Logic in Chemistry, Rouvray, D. (ed.), Academic Press, New York, pp.139-224.

Pettini, M., 2007, Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics, Springer, New York.

Privman, V., M. Fisher, 1983, Finite-size effects at first-order transitions, Journal of Statistical Physics, 33:385-417.

Rosenfeld, A., 1971, Fuzzy groups, Journal of Mathematical Analysis and Applications, 35:512-517.

Stojanovic, M., T. Mitchell, D. Stefanovic, 2002, Deoxyribozyme-based logic gates, Journal of the American Chemical Society, 124:3555-3561.

Tompa, P., C. Szasz L. Buday, 2005, Structural disorder throws new light on moonlighting, Trends in Biochemical Sciences, 30:484-489.

Wallace, R., 2005, Consciousness: A Mathematical Treatment of the Global Neuronal Workspace Model, Springer, New York.

Wallace, R., 2011a, Multifunction moonlighting and intrinsically disordered proteins: information catalysis, non-rigid molecule symmetries and the 'logic gate' spectrum, Comptes Rendus Chimie, 14:1117-1121.

Wallace, R., 2011b, On the evolution of homochirality, Comptes Rendus Biologies, 334:263-268.

Wallace, R., 2012, Spontaneous symmetry breaking in a non-rigid molecule approach to intrinsically disordered proteins, Molecular BioSystems, 8:374-377.

Wallace, R., D. Wallace, 2010, Gene Expression and its Discontents: The Social Production of Chronic Disease, Springer, New York.

Wallace, R., D. Wallace, 2013, A Mathematical Approach to Multilevel, Multiscale Health Interventions: Pharmaceutical industry decline and policy response, Imperial College Press, London.

Weinstein, A., 1996, Groupoids: unifying internal and external symmetry, Notices of the American Mathematical Association, 43:744-752.



PeerJ PrePrints | https://peerj.com/preprints/20v2/ | v2 received: 27 May 2013, published: 28 May 2013, doi: 10.7287/peerj.preprints.20v2