

Version 5

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

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

May 20, 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 the syn- topy group and 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

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 an 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

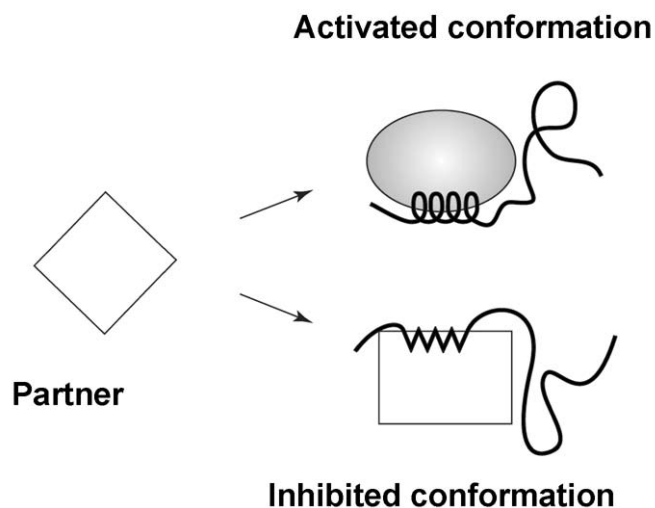


Figure 1: From Tompa et al., 2005. The partner can bind in two ways to the IDP. The top form is activated, and the bottom inhibited. The triggering between the states is done by an ‘information catalysis’ in which an incoming signal shifts the lowest energy state between the two otherwise thermodynamically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

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. Ini-

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

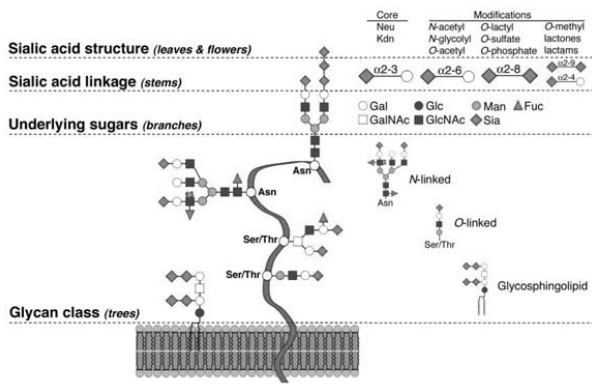


Figure 2: From Cohen and Varki, (2010). Levels of sialome complexity, from core and core modifications to the shifting, bending, twisting, glycan ‘kelp fronds’ that coat most cell surfaces and, via lectin interaction, constitute sophisticated logic gates involved in explosively vast information transfers: 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).

tially, 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.

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.

## 2 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.

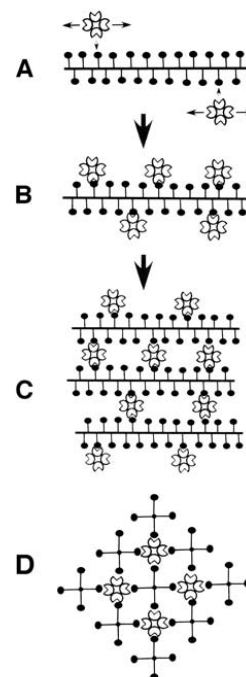


Figure 3: From Dam et al. (2007). Lectin diffuses along and off the flexible glycan frond, until a sufficient number of sites are occupied. Then the coated glycan fronds begin to cross bind, the reaction is complete, and the logic gate is activated. The last figure shows an end view.

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

$$\mathcal{B}_\alpha = C_\beta \mathcal{D}_\gamma \quad (1)$$

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 symmorph groups.

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

### 3 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  $x = [x_1, x_2, \dots]$  and  $y = [y_1, y_2, \dots]$  at times  $i = 1, 2, \dots$ . A joint sequence of signals  $xy \equiv [(x_1, y_1), (x_2, y_2), \dots]$  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.

### 4 Phase transitions and reaction dynamics

Now take the regulatory free energy intensity  $M_Y$  as representing 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). 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 that we do not pursue here (Bos, 2007).

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

$$\mathcal{P}[\omega] = \frac{\exp[-|\chi_\omega|/\kappa\mathcal{T}]}{\sum_\nu \exp[-|\chi_\nu|/\kappa\mathcal{T}]} \quad (5)$$

$\chi_\phi$  is the character of the group element  $\phi$  in that representation, i.e., the trace of the matrix assigned to  $\phi$ , and  $|\dots|$  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

$$\exp[-F/\kappa\mathcal{T}] = \sum_\nu \exp[-|\chi_\nu|/\kappa\mathcal{T}] \quad (6)$$

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, 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 interaction symmetries becomes central to reaction trajectories, according to an Onsager-like nonequilibrium thermodynamics formulation.

Define, then, a ‘symmetry entropy’ based on the Morse Function  $F$  of equation (6) over a set of underlying structural or other parameters  $\mathbf{Q} = [Q_1, \dots, Q_n]$  as the Legendre transform

$$S = F(\mathbf{Q}) - \sum_i Q_i \partial F(\mathbf{Q}) / \partial Q_i. \quad (7)$$

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):

$$dQ_i/dt = \sum_j \mathcal{K}_{i,j} dS/\partial Q_j, \quad (8)$$

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:

$$dQ_t^i = \mathcal{K}_i(t, \mathbf{Q})dt + \sum_j \sigma_{i,j}(t, \mathbf{Q})dB^j, \quad (9)$$

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 (9) 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  $\mathcal{T}$ .

## 5 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

$$S_R = R(D) - DdR/dD = 1/2 \log[\sigma^2/D] + 1/2. \quad (10)$$

The simplest nonequilibrium Onsager equation is just

$$dD/dt = -\mu dS_R/dD = \mu/2D, \quad (11)$$

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

$$D(t) = \sqrt{\mu t}. \quad (12)$$

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 (9), in the presence of an added regulatory system noise indexed by  $\beta/2$ , is

$$dQ_t = [f(Q_t) - G(\mathcal{T})]dt + \frac{\beta}{2} Q_t dW_t, \quad (13)$$

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

$$Q_{equilib} = f^{-1}(G(\mathcal{T})). \quad (14)$$

However, the presence of the noise term can introduce serious complications. Suppose, following the example of equation (11),  $f = \alpha/Q$ . 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

$$dY_t = [2\sqrt{Y_t}(\frac{\alpha}{\sqrt{Y_t}} - G(\mathcal{T})) + \frac{\beta^2}{4} Y_t]dt + \beta Y_t dW_t, \quad (15)$$

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:

$$G(\mathcal{T}) > \beta\sqrt{\alpha/2}. \quad (16)$$

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

Note that if  $f \propto 1/A_n(Q)$ , where  $A_n$  is a polynomial of degree  $n$ , then the equivalent of equation (16) 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

$$\mathcal{T} > G^{-1}(\beta\sqrt{\alpha/2}). \quad (17)$$

Taking a Landau spontaneous symmetry breaking perspective,  $\mathcal{T}$  in equation (17) 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 (9) 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 (9).

## 6 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 (17) 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 associ-

ated with failures of the FLK’s, since these are among the most basic phenomena of the living state.

## 7 References

- Atiyah, M., I. Singer, 1963, The index of elliptical operators on compact manifolds, *Bulletin of the American Mathematical Society*, 69:322-433.
- 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.
- 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.
- 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 submaxillary 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.
- 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.
- 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.
- Longuet-Higgins, H., 1963, The symmetry groups of non-rigid molecules, *Molecular Physics*, 6:445-460.
- Maturana, H., 1970, The biology of cognition, BCL Research Report 9.0, University of Illinois.
- 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.
- Tompa, P., C. Szasz L. Buday, 2005, Structural disorder throws new light on moonlighting, *Trends in Biochemical Sciences*, 30:484-489.

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.