Version 8.1

Beyond the fuzzy lock-and-key: spontaneous symmetry shifts and glycan/lectin logic gates

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

January 30, 2012

Abstract

Changes in the molecular topology of glycan/lectin interaction may explain observed reaction punctuation driven by experimental gradients in reactant concentration. Adoption of a 'biological renormalization' perspective from statistical physics for the analysis of such phase transitions suggests, in marked contrast to conventional physical systems, a broad spectrum of possible universality class behaviors. This spectrum may, in typical perverse biological manner, be of central scientific interest. Generalization, via formalism abducted from coevolutionary theory, suggests that glycan/lectin molecular switches instantiate logic gates that may be as sophisticated as those characterizing basic neural process, if on a different scale.

Keywords: free energy, information theory, renormalization, symmetry breaking

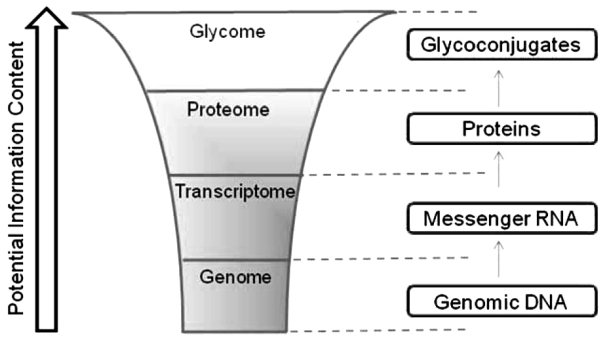

Figure 1: From Gupta et al. (2010). There is an exponential increase in potential information content from the genome through the proteome up to the glycome. Since information is a form of free energy, it is possible to adapt the usual renormalization treatment of phase changes, taking an analog to Wilson's (1971) approach.

1 Introduction

Figure 1, from Gupta et al. (2010), illustrates the exponential increase in the amount of potential information content from the genome through the proteome up to the glycome. The 12 basic mammalian monosaccharides form branches with sidechains that may represent 7 to 10 thousand possible 'glycome determinants' (Cummings, 2009; Wallace, 2012c), forming larger structures that interact with lectin proteins to actually transmit biological information.

In vitro, punctuated transitions in glycan/lectin interaction topology cause large-scale phase change in reactions driven by experimental gradients in reactant concentration (Dam and Brewer, 2010). Although a comprehensive theory probably requires a treatment of reaction funnel dynamics analogous to recent studies of protein folding (e.g., Wallace, 2011), it is possible to adapt a phenomenological 'renormalization' strategy from statistical physics to such phase transitions. The essence of the argument is that information is a form of free energy, so that something resembling Wilson's (1971) renormalization methods can be abducted to complicated biological phase changes associated with biological information transmission.

The transmission of information, however, inevitably involves a source that actually generates that information, something that 'speaks' a 'language', and such processes are constrained by the necessary conditions of the asymptotic limit theorems of information theory: the Shannon Coding Theorem, the Shannon-McMillan Source Coding Theorem, and the Rate Distortion Theorem, and variants like the information theory chain rule (Ash 1990; Cover and Thomas 2006; Khinchin 1957).

2 The critical exponent

For this work, the essential matter is the homology between information source uncertainty – the richness of the language being spoken – and the free energy density of a physical sys-

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

tem. This allows adaptation of Wilson's (1971) renormalization symmetry methods for phase transitions.

Given an information source that produces as sequence of signals having structure – loosely, grammar and syntax – the Shannon-McMillan Theorem says that such utterances can be broken into two disjoint pieces, a very large set of gibberish that has vanishingly low probability, and a very small set in accordance with the rules of grammar and syntax characterizing the information source for which the following condition holds:

Let $N(n)$ be the number of grammatical and syntactical statements of length n produced by an information source X. Then the limit

$$H[X] = \lim_{n\to\infty} \frac{\log[N(n)]}{n} \tag{1}$$

exists and is independent of the statement itself. That is, H is a universal constant for the information source. If the limit converges for some finite n_0, than that number is called the order of the information source.

A second limiting relation is that the statements produced by any information source must be transmitted along a channel having a channel capacity $C \geq 0$ such that

$$H[X] \leq C. \tag{2}$$

Details can be found in any number of texts (Ash, 1990; Cover and Thomas, 2006; Khinchin, 1957).

The free energy density of a physical system having volume V and partition function $Z(\beta, V)$ derived from the system's Hamiltonian – the energy function – at temperature β is (e.g., Landau and Lifshitz 2007)

$$F[\beta] = \lim_{V\to\infty} -\beta\frac{\log[Z(\beta,V)]}{V} \equiv \lim_{V\to\infty} \frac{\log[\hat{Z}(\beta,V)]}{V}, \tag{3}$$

with $\hat{Z} = Z^{-\beta}$. The latter expression is formally similar to equation (1), 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, 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 inevitable – and massive – losses via the second law of thermodynamics.

Here we will use Wilson's renormalization strategy to characterize the behavior of channel capacity C near critical values of driving parameters. The basic argument is well known, and the Wikipedia entry on critical exponents is as good a source as any:

Above and below the critical value of some driving parameter, say Tc, the system of interest has two distinct phases, characterized by an order parameter that vanishes above Tc. Here we will take an empirical index of channel capacity as the order parameter. Let

$$\tau \equiv \frac{T - Tc}{Tc}.$$

We are interested in the first term of a series expansion of C in tau,

$$C(\tau) = A\tau^{\alpha}(1 + b\tau^{\alpha_1}...),$$

so that, in first order,

$$C(\tau) \propto \tau^{\alpha}.$$

A simple calculation gives

$$\alpha = \lim_{\tau\to 0} \frac{\log|C(\tau)|}{\log|\tau|}. \tag{4}$$

The hard trick is to calculate α from first principles. For (relatively) simple physical phenomena such exponents are universal across many systems, a function of simple underlying renormalization symmetry relations (Wilson, 1971; Binney et al., 1992). A detailed calculation in the Supplemental Material, however, shows that for biological structures, a vast array of 'biological' renormalizations are possible, and universality classes – collections of phenomena having the same α – may be limited to sets of very similar reacting species.

H and C, as free energy measures, can also be viewed as Morse Functions, in the sense of Pettini (2007), and thus subject to the topological hypothesis: Singularities in these measures – critical points – are to be associated with a fundamental change in underlying topology of the manifold on

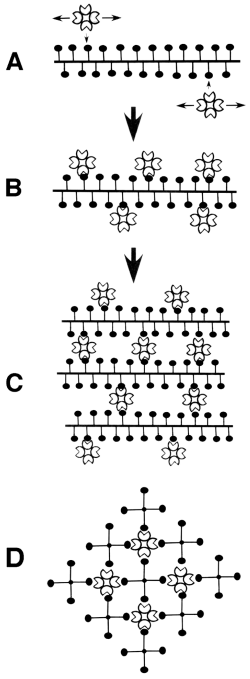

Figure 2: From Dam et al. (2007). (A) At first, lectin diffuses along (and off) the glycan kelp frond, until, (B), a sufficient number of sites are occupied. Then (C), the lectin-coated glycan fronds begin to cross bind and the reaction is saturated. (D) shows an end-on view of the complex in (C).

which these measures are defined. This is a generalization of Landau's observation (Landau and Lifshitz, 2007) that second order phase transitions in simple physical systems, those without latent energy, are usually characterized by changes in underlying symmetry, with the higher energy states being more symmetric. We now impose this perspective, albeit, as will become evident, in an inverse manner.

3 Two examples

3.1 Area concentration

The carbohydrate α-GalNAc interacts with the lectin biotinylated soybean agglutinin (SBA) in solution to form a sequence of increasingly complicated interlinked conformations at appropriate concentrations of reacting species. Dam et al. (2007) describe this 'bind-and-slide' process in terms of a change in topology, according to figure 2.

Initially, the lectin diffuses along (and off) the glycan kelp frond, until a number of sites are occupied. Then the lectin-coated glycan fronds begin to cross bind, until the reaction saturates. Figure 2D shows an end-on view of the complex shown longitudinally in 2C.

Dam and Brewer (2008) generalize this as follows:

> The bind and slide model for lectins binding to multivalent glycosides, globular, and linear glycoproteins is distinct from the classical 'lock and key' model for ligand-receptor interactions. The bind and slide (internal diffusion) model allows a small fraction of bound lectin molecules to dynamically move from carbohydrate to carbohydrate epitope in globular and linear glycoproteins. This, in turn, can facilitate lectin-mediated cross-linking of such glycoproteins on the surface of cells... Such cross-linked receptors, in turn, trigger signal transduction mechanisms... Indeed, a large number of transmembrane receptors are found clustered... Thus the affinity and hence specificity of ligand-receptor interactions may be regulated by epitope and receptor clustering in many biological systems.

Under typical physiological circumstances, glycans form a literal kelp bed bound to cellular surfaces, and the essential parameter becomes area density of the fronds. The excellent review article of Dam and Brewer (2010) describes the work of Oyelaran et al. (2009), who conducted a series of heroic density-dependent fluorescence experiments, and it becomes possible to take the observed intensity of that fluorescence as an index of channel capacity, since no information transmission $\rightarrow$ no reaction $\rightarrow$ no fluorescence.

Using microarray methods, Oyelaran et al. embedded α-GalNAc onto bovine serum albumin (BSA), with different numbers of carbohydrate molecules (CM) per BSA site, typically ranging from 4 to 40. At density of 4, the CM were separated by about 85 Angstroms, and at 20, by about 40 Angstroms.

Typically, fluorescence intensity, K, under such conditions, follows a relation

$$K = \frac{Km}{Kd/L + 1} \tag{5}$$

where Km the maximum intensity, Kd the apparent dissociation constant for interaction between protein and immobilized glycoconjugate, and L the concentration of lectin. See figure 3.

The essential matter is the way in which Kd varies with glycan area density. For α-GalNAc embedded on BSA, again reacting with biotinylated soybean agglutinin, SBA, Kd followed the relation of figure 4.

Kd in equation (5) almost disappears with increasing number of glycan molecules per BSA site: The value falls from near 4200 at n=4 to 194 at n=9, so that Kd is, from a physics perspective, an (inverse) 'order parameter' that undergoes a change representing a topological transformation, here a cross-linking phase transition.

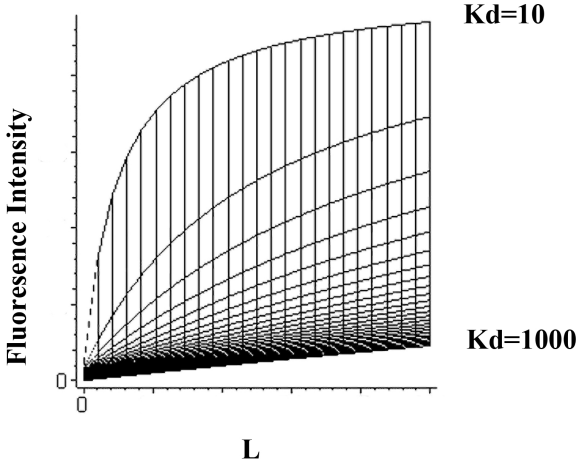

Figure 3: Fluorescence intensity vs. lectin concentration for different values of Kd. Large Kd implies little fluorescence, and hence, by our arguments, small channel capacity, C.

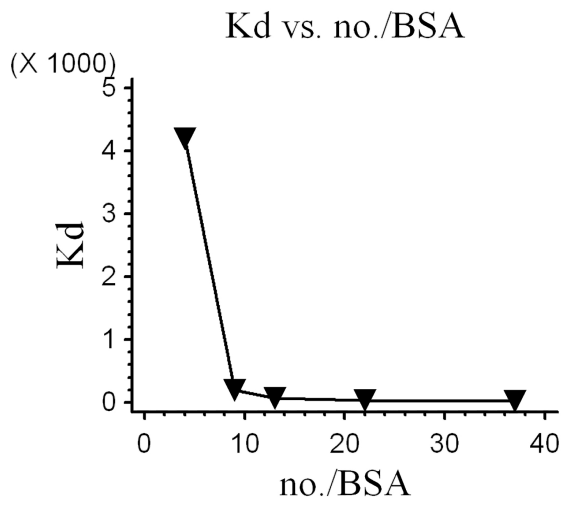

Figure 4: Kd vs. no. glycan molecules per BSA site. The value drops precipitously from about 4200 at n=4 to 194 at n=9, so that Kd is an (inverse) order parameter that undergoes a change representing a topological transformation, here a cross-linking phase transition. A finer grain analysis would permit estimation of α in the relation 'Order Parameter' $\propto [(\rho - \rho_C)/\rho_C]^\alpha$ near the critical density ρ_C. Calculating an α from first principles would be a considerable scientific achievement.

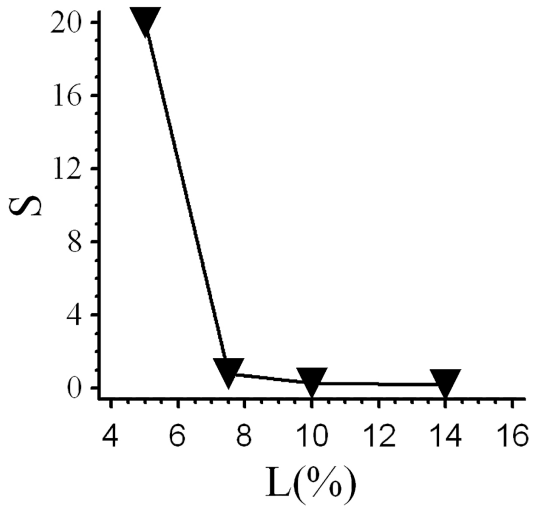

Figure 5: From Orr et al. (1979). Critical concentration effect in aggregation of lecithin liposomes by fixed concentration of ConA. S, the inverse pseudo rate constant, is the (inverse) order parameter that disappears at higher concentrations of lectin.

3.2 Volume concentration

As described by by Dam and Brewer (2010), classic work by Orr et al. (1979) examined aggregation of lecithin liposomes having a synthetic mannose glycolipid by concanavalin A (ConA) at 100gm/ml. The mol concentration of the glycolipid in the liposomes ranged from 5 to 14 percent, while the characteristic reaction duration – again, a kind of inverse order parameter – varied from 0.2 to 20.0 sec., as in figure 5.

Orr et al. (1979) state

> It is interesting to note that the lectin-induced aggregation exhibits a threshold or critical concentration effect. At incorporation levels of 5 mol % and less, the rate of the absorbance increase is slow, whereas at 7.5 mol % a dramatic increase in the rate is observed.

In both experiments, finer grain observations would allow determination of α in the relation 'Order Parameter' $\propto [(\rho - \rho_C)/\rho_C]^\alpha$ near the critical area or volume density ρ_C. Calculating an α from first principles for different reacting chemical species would be a scientific tour de force.

4 Information catalysis

Information per se does not carry very much free energy, but the mechanisms that instantiate signals do, and this fact, in concert with the asymptotic limit theorems of information

theory, seems to permit an important evolutionary exaptation of entropic loss. Here, the essential parameter of interest may be density measures, but the argument seems more general.

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):

$$H_{X,Y} < H_X + H_Y. \tag{6}$$

Typically, one might expect the average production of information, $\hat{H}$, from a process having an available metabolic free energy rate M, to follow a relation of the form

$$\hat{H} = \frac{\int H \exp[-H/\kappa M] dH}{\int \exp[-H/\kappa M] dH} \approx \kappa M, \tag{7}$$

where κ is quite small, so the integral converges.

Then, from the chain rule,

$$\hat{H}_{X,Y} < \hat{H}_X + \hat{H}_Y,$$

$$M_{X,Y} < M_X + M_Y. \tag{8}$$

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. Thus high entropic loss – small κ – becomes a tool for regulating biochemical reactions.

5 A 'coevolutionary' model

Assume a larger set of interacting information sources, H_j, each associated with a free energy intensity M_j. We first write each information source $\hat{H}_j \propto M_j$ as a function of a vector of 'density-like' parameters $\mathbf{K} = [K_1, K_2, ...]$.

For a simple physical system, one would expect a nonequilibrium thermodynamics driven by gradients in an entropy-like factor constructed from the $\hat{H}_j$ via analogs to empirical Onsager relations.

Define a set of 'information entropies'

$$S_j \equiv \hat{H}_j - \sum_q K_q \partial \hat{H}_j / \partial K_q \propto$$

$$M_j - \sum_q K_q \partial M_j / \partial K_q. \tag{9}$$

The simplest Onsager-like approach imposes dynamics driven by the gradients of the entropies as

$$dK_i/dt = \sum_j L_{i,j} \partial S_j / \partial K_j. \tag{10}$$

However, following Champagnat et al. (2006), it is possible to take quite a different perspective, a 'coevolutionary' stochastic generalization in which the 'parameter vector' is the set of 'information' variates itself, $\mathbf{Q} = [\hat{H}_1, \hat{H}_2, ...] \propto [M_1, M_2, ...]$ rather than a set of external driving terms. $\mathbf{Q}$ would, in turn, be constrained by a vector of channel capacities. Then, as in Champagnat et al., the dynamical relations are given by phenomenological stochastic differential equations of the form

$$dQ_t^i = L^i(t, \mathbf{Q}) dt + \sum_j \sigma^{i,j}(t, \mathbf{Q}) dB^j, \tag{11}$$

or an analog using channel capacities. L and σ are appropriate functions and the dB represent stochastic 'noise' having characteristic quadratic variation (e.g., Protter, 1990).

These stochastic differential equations produce a coevolutionary – mutually driving – system so that setting them to to zero generates a collection of quasi-stable equilibria where transitions between them will be driven by 'large deviations' associated with yet other information sources: larger-scale, embedding, regulatory systems. The calculation is classic, and such large deviations can usually be represented by some entropy-like measure (e.g., Dembo and Zeitouni, 1998) that, in a biological context, is an embedding information source:

For figures 4 and 5, there are two such quasi-equilibria, described by the two phases indexed by collapse of the inverse order parameters, with experimenter-driven information – the systematic 'large deviation' of experimental changes in density measures – triggering the transition between them. These results might be viewed as examples of biological logic gates, significantly larger, of course, than the more familiar neural synapse.

6 Discussion and conclusions

For even a 'simple' renormalization model of glycan/lectin reaction phase transitions, the 'universality' exponent α, following the arguments of the Supplemental Material, is unlikely to be universal, and indeed should be a precisely distinguishing characteristic of both the reacting species and the modes of their reactions. Thus the 'failure' of universality might well be a tool that allows insight into the energy landscapes of $M_{X,Y}, M_X$ and M_Y, the resulting canalization via information catalysis, and its 'coevolutionary' correlates. Indeed, biology is not physics, and, typically, evolution may exapt, in the sense of Gould (1992), mechanisms that can distinguish reacting chemical species and/or their modes of reaction.

Underlying the general approach is a biological version of the spontaneous symmetry change perspective so popular in current physical theory. The symmetries, however, may not be as 'pure' as those most familiar to physicists and, as Wallace (2012b) argues, might involve such complexities as groupoid tilings and their wreath products. Nonetheless, the line of argument implied by figures 2 and 4 is quite compelling, and the techniques of the Supplemental Materials may aid in the calculation of spectra of 'universality' constants for such experiments from first principles.

What becomes obvious, almost in passing, is the utterly central point that the *in vitro* glycan/lectin phase transitions characterized by figures 4 and 5 imply the operation of complicated biological logic gates *in vivo* that must be nearly as sophisticated as the more familiar neural synaptic switches, if on different scales. Cascades of even 'simple' logic gates can carry out very complex computational processes, and these arguments add weight to an emerging perspective that sees the living state as characterized by cognitive phenomena at virtually every scale and level of organization.

7 Acknowledgments

The author thanks Drs. D.N. Wallace and C.F. Brewer for useful discussions. The expressed opinions, however, are those of the author.

8 References

Ash, R., 1990, Information theory, Dover, New York.

Binney, J., N. Dowrick, A. Fisher, M. Newman, 1992, The theory of critical phenomena, Clarendon Press, Oxford.

Champagnat, N., R. Ferriere, S. Meleard, 2006, Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models, Theoretical Population Biology, 69:297-321.

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, F. Brewer, 2007, Binding studies of α-GalNAc-specific lectins to the α-GalNAc(Tn-antigen) form of porcine submaxilary mucin and its smaller fragments, Journal of Biological Chemistry, 38:28256-28263.

Dam, T., C.F. Brewer, 2008, Effects of clustered epitopes in multivalent ligand-receptor interactions, Biochemistry, 47:8470-8476.

Dam, T., C.F. Brewer, 2010, Lectins as pattern recognition molecules: The effects of epitope density in innate immunity, Glycobiology, 20:270-279.

Dembo, A., O. Zeitouni, 1998, Large deviations: techniques and applications, 2nd edition, Springer, New York.

Feynman, R., 2000, Feynman lectures on computation, Westview Press.

Gould, S., 1992, The structure of evolutionary theory, Harvard University Press, Cambridge, MA.

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

Khinchin, A., 1957, The mathematical foundations of information theory, Dover Publications, New York.

Landau, L., E. Lifshitz, 2007, Statistical physics, 3rd edition, Part 1, Elsevier, New York.

Orr, G., R. Rando, F. Wen Bangerter, 1979, Synthetic glycolipids and the lectin-mediated aggregation of liposomes, Journal of Biological Chemistry, 254:4721-4725.

Oyelaran, O., Q. Li, D. Farnsworth, J. Gildersleeve, 2009, Microarrays with varying carbohydrate density reveal distinct subpopulations of serum antibodies, Journal of Proteome Research, 8:3529-3538.

Pettini, M., 2007, Geometry and topology in Hamiltonian dynamics and statistical mechanics, Springer, New York.

Protter, P., 1990, Stochastic integration and differential equations: A new approach, Springer, New York.

Wallace, R., 2011, Structure and dynamics of the 'protein folding code' inferred using Tlusty's topological rate distortion approach, BioSystems, 103:18-26.

Wallace, R., 2012a, Empirical predictions of an intrinsically disordered protein theory approach to glycan/lectin interaction kinetics,
http://precedings.nature.com/documents/6752/version/1
Wallace, R., 2012b, Spontaneous symmetry breaking in a non-rigid molecule approach to intrinsically disordered proteins, Molecular BioSystems, 8:374-377.
Wallace, R., 2012c, Extending Tlusty's rate distortion index theorem method to the glycome: do even 'low level' biochemical phenomena require sophisticated cognitive paradigms? In press, BioSystems.
Wilson, K., 1971, Renormalization group and critical phenomena. I Renormalization group and the Kadanoff scaling picture, Physical Review B, 4:3174-3183.

9 Supplemental material – 'biological' renormalization

Following the classic phase transition arguments of Wilson (1971), the quantity of interest, F, and the correlation length – the degree of coherence along the embedding structure of interest – scale under renormalization clustering in chunks of size R, here taken as an appropriate topological distance measure, as

$$F[Q_R, J_R]/f(R) = F[J, Q]$$

$$\chi[Q_R, J_R]R = \chi(Q, J),$$

(12)

with $f(1) = 1, Q_1 = Q, J_1 = J$. Q is to be seen as an inverse temperature analog, and in the limit, again following the patterning of Wilson (1971), we will allow the 'external field' $J \to 0$. Other approaches are possible.

Differentiating these two equations with respect to R, so that the right hand sides are zero, and solving for dQ_R/dR and dJ_R/dR gives, after some consolidation, expressions of the form

$$dQ_R/dR = u_1 d\log(f)/dR + u_2/R$$

$$dJ_R/dR = v_1 J_R d\log(f)/dR + \frac{v_2}{R} J_R.$$

(13)

The $u_i, v_i, i = 1, 2$ are functions of Q_R, J_R, but not explicitly of R itself.

We expand these equations about the *critical value* $Q_R = Q_C$ and about $J_R = 0$, obtaining

$$dQ_R/dR = (Q_R - Q_C)y d\log(f)/dR + (Q_R - Q_C)z/R$$

$$dJ_R/dR = wJ_R d\log(f)/dR + xJ_R/R.$$

(14)

The terms $y = du_1/dQ_R|_{Q_R=Q_C}, z = du_2/dQ_R|_{Q_R=Q_C}, w = v_1(Q_C, 0), x = v_2(Q_C, 0)$ are constants.

Solving the first of these equations gives

$$Q_R = Q_C + (Q - Q_C)R^z f(R)^y,$$

(15)

again remembering that $Q_1 = Q, J_1 = J, f(1) = 1$.

Wilson's essential trick is to iterate on this relation, which is supposed to converge rapidly near the critical point (Binney et al., 1992), assuming that for Q_R near Q_C, we have

$$Q_C/2 \approx Q_C + (Q - Q_C)R^z f(R)^y.$$

(16)

We iterate in two steps, first solving this for $f(R)$ in terms of known values, and then solving for R, finding a value R_C that we then substitute into the first of equations (12) to obtain an expression for $H[Q, 0]$ in terms of known functions and parameter values.

The first step gives the general result

$$f(R_C) \approx \frac{[Q_C/(Q_C - Q)]^{1/y}}{2^{1/y} R_C^{z/y}}.$$

(17)

Solving this for R_C and substituting into the first expression of equation (12) gives, as a first iteration of a far more general procedure (Shirkov and Kovalev, 2001), the result

$$F[Q,0] \approx \frac{F[Q_C/2,0]}{f(R_C)} = \frac{F_0}{f(R_C)}$$

$$\chi(Q,0) \approx \chi(Q_C/2,0)R_C = \chi_0 R_C,$$

(18)

which are the essential relationships.

Note that a power law of the form $f(R) = R^m, m = 3$, which is the direct physical analog, may not be biologically reasonable, since it says that F can grow very rapidly as a function of increased substrait size. Such rapid growth is not necessarily observed.

Taking the biologically realistic example of non-integral 'fractal' exponential growth,

$$f(R) = R^\delta,$$

(19)

where $\delta > 0$ is a real number which may be quite small, equation (17) can be solved for R_C, obtaining

$$R_C = \frac{[Q_C/(Q_C - Q)]^{[1/(\delta y+z)]}}{2^{1/(\delta y+z)}}$$

(20)

for Q near Q_C. Note that, for a given value of y, one might characterize the relation $\alpha \equiv \delta y + z =$ constant as a 'tunable universality class relation' in the sense of Albert and Barabasi (2002).

Substituting this value for R_C back into equation (17) gives a complex expression for F, having three parameters: δ, y, z.

A more biologically interesting choice for $f(R)$ is a logarithmic curve that 'tops out', for example

$$f(R) = m\log(R) + 1.$$

(21)

Again $f(1) = 1$.

Using Mathematica 4.2 or above to solve equation (17) for R_C gives

$$R_C = [\frac{S}{LambertW[S\exp(z/my)]}]^{y/z},$$

(22)

where

$$S \equiv (z/my)2^{-1/y}[Q_C/(Q_C - Q)]^{1/y}.$$

The transcendental function LambertW(x) is defined by the relation

$$LambertW(x)\exp(LambertW(x)) = x.$$

It arises in the theory of random networks and in renormalization strategies for quantum field theories.

An asymptotic relation for $f(R)$ would be of particular biological interest, implying that F increases to a limiting value with population growth. Such a pattern is broadly consistent with, for example, calculations of the degree of allelic heterozygosity as a function of population size under a balance between genetic drift and neutral mutation (Hartl and Clark, 1997; Ridley, 1996). Taking

$$f(R) = \exp[m(R-1)/R]$$

(23)

gives a system which begins at 1 when $R = 1$, and approaches the asymptotic limit $\exp(m)$ as $R \to \infty$. Mathematica 4.2 finds

$$R_C = \frac{my/z}{LambertW[A]},$$

(24)

where

$$A \equiv (my/z)\exp(my/z)[2^{1/y}[Q_C/(Q_C - Q)]^{-1/y}]^{y/z}.$$

These developments indicate the possibility of taking the theory significantly beyond arguments by abduction from simple physical models, although the notorious difficulty of implementing information theory existence arguments will undoubtedly persist.

9.1 Universality class

Physical systems undergoing phase transition usually have relatively pure renormalization properties, with quite different systems clumped into the same 'universality class,' having fixed exponents at transition (Binney et al., 1986). Biological phenomena may be far more complicated:

If the system of interest is a mix of subgroups with different values of some significant renormalization parameter m in the expression for $f(R, m)$, according to a distribution $\rho(m)$, then, at least to first order,

$$F[Q_R, J_R] =< f(R, m) > F[Q, J]$$

$$\equiv F[Q, J] \int f(R, m)\rho(m)dm. \tag{25}$$

If $f(R) = 1 + m\log(R)$ then, given any distribution for m,

$$< f(R) >= 1 + < m > \log(R) \tag{26}$$

where $< m >$ is simply the mean of m over that distribution.

Other forms of $f(R)$ having more complicated dependencies on the distributed parameter or parameters, like the power law R^δ, do not produce such a simple result. Taking $\rho(\delta)$ as a normal distribution, for example, gives

$$< R^\delta >= R^{<\delta>} \exp[(1/2)(\log(R^\sigma))^2], \tag{27}$$

where σ^2 is the distribution variance. The renormalization properties of this function can be determined from equation (12), and the calculation is left as an exercise, best done in Mathematica 4.2 or above.

Thus the phase transition properties of mixed systems will not in general be simply related to those of a single subcomponent, a matter of possible empirical importance: If sets of relevant parameters defining renormalization universality classes are indeed distributed, experiments observing pure phase changes may be very difficult. Tuning among different possible renormalization strategies in response to external signals would result in even greater ambiguity in recognizing and classifying biological phase transitions.

Important aspects of mechanism may be reflected in the combination of renormalization properties and the details of their distribution across subsystems.

In sum, biological systems are likely to have very rich patterns of phase transition that may not display the simplistic, indeed, literally elemental, purity familiar to physicists. Overall mechanisms will, however, still remain significantly constrained by the theory, in the general sense of probability limit theorems.

The more biologically realistic renormalization strategies given above produce sets of several parameters defining the universality class.

Suppose, now, that the renormalization properties of a system at some 'time' k are characterized by a set of (possibly coarse-grained) parameters $A_k \equiv \alpha_1^k, ..., \alpha_m^k$. Fixed parameter values define a particular universality class for the renormalization. We suppose that, over a sequence of 'times,' the universality class properties can be characterized by a path $x_n = A_0, A_1, ..., A_{n-1}$ having significant serial correlations which, in fact, permit definition of an adiabatically piecewise stationary ergodic information source associated with the paths x_n. We call that source $\mathbf{X}$.

Suppose also that the set of impinging signals is also highly structured and forms another information source $\mathbf{Y}$ that interacts not only with the system of interest globally, but specifically with its universality class properties as characterized by $\mathbf{X}$. $\mathbf{Y}$ is necessarily associated with a set of paths y_n.

Pair the two sets of paths into a joint path, $z_n \equiv (x_n, y_y)$ and invoke an inverse coupling parameter, Q, between the information sources and their paths. This leads, by the arguments above, to phase transition punctuation of $I[Q]$, the mutual information between $\mathbf{X}$ and $\mathbf{Y}$, under either the Joint Asymptotic Equipartition Theorem or under limitation by a distortion measure, through the Rate Distortion Theorem. The essential point is that $I[Q]$ is a splitting criterion under these theorems, and thus partakes of the homology with free energy density.

Activation of universality class tuning then becomes itself a punctuated event in response to increasing linkage between the biological structure of interest and an external structured signal or some particular system of internal events.

This iterated argument exactly parallels the extension of the General Linear Model to the Hierarchical Linear Model in regression theory (Byrk and Raudenbusch, 2001).

10 Supplemental references

Albert, R., A. Barabasi, 2002, Statistical mechanics of complex networks, Reviews of Modern Physics, 74:47-97.

Binney, J., N. Dowrick, A. Fisher, M. Newman, 1992, The theory of critical phenomena, Clarendon Press, Oxford.

Byrk, A., S. Raudenbusck, 2001, Hierarchical linear models: applications and data analysis methods, Sage Publications, New York.

Hartl, D., A. Clark, 1997, Principles of population genetics, Sinaur Associates, Sunderland, MA.

Ridley, M., 1996, Evolution, Second Edition, Blackwell Science, Oxford University Press.

Shirkov, D., V. Kovalev, 2001, The Bogoliubov renormalization group and solution symmetry in mathematical physics, Physics Reports, 352:219-249.

Wilson, K., 1971, Renormalization group and critical phenomena. I Renormalization group and the Kadanoff scaling picture, Physical Review B, 4:3174-3183.