Version 7

Empirical predictions of an intrinsically disordered protein theory approach to glycan/lectin reaction kinetics

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

January 2, 2012

Abstract

Newly-developed methods from the theory of intrinsically disordered proteins can be applied to the flexible glycan structures that coat cellular surfaces and provide rich channels for biological information transmission. Extension of a mechanistic 'arm-in-sleeve' model via a nonrigid molecule symmetry analysis leads to expectation of empirical observation of punctuated 'spectral' classifications in glycan/lectin interaction parameterized by an appropriate index of glycan frond length or other index of topological complexity, possibly requiring groupoid classifications analogous to quasicrystals.

Key Words: group theory, groupoid, information theory, nonrigid molecule, rate distortion

1 Introduction

Flexible glycan fronds coat the cellular surface and provide a signaling base for a vast spectrum of interactions with other biological structures and entities, ranging from tissue partners to pathogens and parasites. The are built from some 7-10 thousand 'glycan determinants', in the sense of Cummings (2009), making glycan structures far more complicated than proteins constructed from 20 or so amino acids, intractable as understanding both protein folding and intrinsically disordered protein (IDP) function remain.

Wallace (2012a) has examined the reducto ad absurdum implied by application of Tlusty's elegant rate distortion error code analysis in the context of so many fundamental building blocks, finding that the production of flexible glycan fronds at the cell surface must be regulated by sophisticated processes of chemical cognition. Here we focus down on the business end of the matter, where the glycan meets the lectin.

As is well known, while evolutionary conservation is roughly in the order of genetic code > RNA sequences > primary protein sequence > metabolic pathways > cellular lipid composition > surface glycan structures, the information content, related to structural diversity, is in reverse order.

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

Most glycans, at the information extreme of this series, have significant freedom of motion in aqueous solution. In addition, most intrinsic glycan functions are further mediated, not by a single absolutely required sequence, but by an ensemble of possible glycan conformations, spanning a continuum that can carry out functions.

The essential chemical interaction involves an anchored glycan surface module and an incoming lectin or other chemical sensor, a signal exchange that Gabius et al. (2002) characterize in terms of a 'sugar code'.

Glycan flexibility may serve an essential purpose for reaction kinetics, analogous to the increased reaction rates available to IDP (Gruebele, 2005). The argument adapts recent formal results in IDP theory (Wallace, 2011a, b, 2012b) to the analysis of glycan/lectin interactions.

2 Catalysis by mechanical flexibility

The first step is to characterize the reaction between glycan and lectin in terms of a sequence of intermediate states in an analog to the free energy funnel of protein folding (Gruebele, 2005), essentially an arm-in-sleeve fitting mechanism, and we are interested in the sequence of intermediate states leading to the final fit. The first – plausible – assumption is that, starting with an initial contact state a_0, there is a subsequent identifiable series of glycan/lectin conformations $x_n \equiv \{a_0, a_1, ..., a_n\}$ having the property that a small set is highly probable, but the vast majority will be in violation of an inherent conformational chemical grammar and syntax. The formal requirement is that, given $N(n)$ highly probable paths of length n, the limit

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

exists and is independent of path. Under such a circumstance it is possible to define joint and conditional probabilities $P(a_0, ..., a_n), P(a_n|a_0, ..., a_{n-1})$ such that the appropriate joint and conditional Shannon uncertainties of the form $-\sum_i P_i \log[P_i]$ exist, and the underlying conditions of the Shannon-McMillan and Rate Distortion Theorems apply. See Wallace (2011a) for details. It is also possible to define an average measure of distortion, say D, between an intermediate glycan/lectin conformation and the final desired state. According to the Rate Distortion Theorem (Cover and Thomas, 2006), there is a minimum channel capacity, R, such that if the 'reaction message' x is transmitted at a rate less than or equal to R, then the average distortion of the final configuration will be less than D, for any reasonable measure of distortion. In addition, the relation between R and D will always be convex, i.e., a reverse-J shaped curve.

For a Gaussian channel having noise of zero mean and variance σ^2 (Cover and Thomas, 2006),

$$R(D) = (1/2)\log[\sigma^2/D], D < \sigma^2$$

$$R(D) = 0, D \geq \sigma^2.$$

(2)

As Feynman (2000) argues, information is a form of free energy, and it is easy to construct an ideal machine that can turn the information within a message into useful work, the essential characteristic of free energy. Arguing by abduction from intrinsically disordered protein theory (Wallace, 2011a), we suppose the channel capacity of the glycan/lectin sleeve mechanism is capped at some maximum possible value R, and that the glycan/lectin interaction of equation (1) is characterized by some index of topological complexity, λ. That is, we have a set of limits of the form H_λ depending on interaction structure. What, under such circumstances, will be the reaction rate of the glycan/lectin interaction? Following the intrinsically disordered protein argument of Wallace (2011a), under limited channel capacity, the probability of H_λ – defining the reaction rate – will be given by something of the form

$$P[H_\lambda] = \frac{\exp[-H_\lambda/\mu R]}{\sum_\beta \exp[-H_\beta/\mu R]},$$

(3)

where μ is a scaling constant and λ, β index topological complexity, so that

$$\log[P(H_\lambda)] = \log[\frac{\exp[-H_\lambda/\mu R]}{\sum_\beta \exp[-H_\beta/\mu R]}] \equiv C(R) - \frac{H_\lambda}{\mu R}.$$

(4)

The simplest assumption is that $H_\lambda \propto \lambda$. Then, using an integral approximation for the sum,

$$P[\lambda] = \frac{\exp[-m\lambda/\mu R]}{\int_{\beta=0}^{\infty} \exp[-m\beta/\mu R]} = \frac{m}{\mu R}\exp[-\frac{m\lambda}{\mu R}]$$

(5)

and

$$\log[P(\lambda)] = \log[m/\mu R] - \frac{m\lambda}{\mu R}.$$

(6)

This is closely similar to the classic result for protein folding rate vs. absolute contact order, an index of protein topological complexity (Gruebele, 2005).

Assuming a Gaussian channel, we can nest equation (2) within this result to obtain a spectrum of reaction rate relations that now depend on both the topological complexity of the glycan/lectin sleeve-fitting reaction and on the channel noise index σ^2, as in figure 1.

Consider λ fixed by the chemistry of the glycan/lectin interaction. Following the arguments of Wallace (2011a), the flexibility of the glycan frond acts as a self-lubricating topological catalyst that decreases σ^2, increasing the reaction rate with incoming species above what would be expected from a rigid molecular structure. Something similar might emerge from equation (4) directly, without need to invoke simplifying assumptions of linear dependence of topological interaction complexity through the index λ that is, in the context of protein folding, the absolute contact order. Indeed, by such mechanisms, glycan flexibility may define its attractiveness to pathogens, suggesting another evolutionary necessity for rapid changes in glycan structure.

Next, we derive the self-lubrication result using formal methods, finding the possibility of an observable spectral decomposition for glycan/lectin interaction.

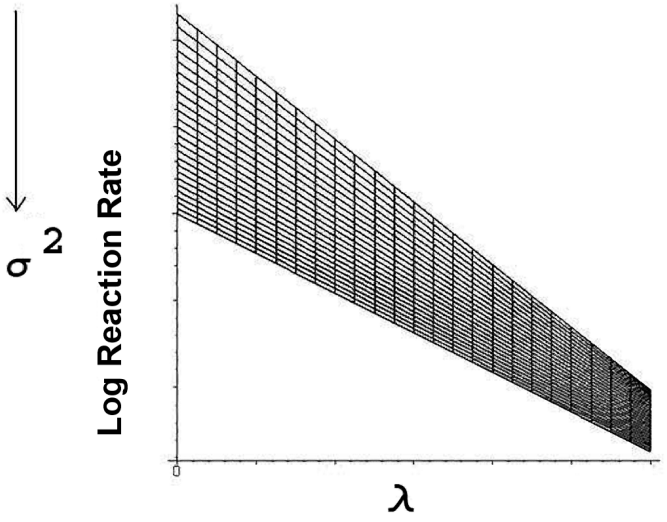

Figure 1: Spectrum of relations between log reaction rate and increasing topological complexity of glycan/lectin interaction for increasing 'roughness' of the 'reaction rate funnel' connecting glycan determinants with incoming signal species as measured by the noise σ^2 for a Gaussian channel. λ increases to the right and σ^2 increases downward, lowering reaction rates. Glycan flexibility acts as a self-lubricating topological catalyst that decreases σ^2, increasing the reaction rate with incoming species above what would be expected from a rigid molecular structure.

3 Nonrigid molecule theory

The line of argument can be made more precise by invoking same kind of nonrigid molecule argument that Wallace (2011b, 2012b) has applied to intrinsically disordered proteins: Although we like to think of IDP and glycan/lectin reactions in classical terms, invoking such metaphors as 'fly-casting' or 'a snake slithering', these are actually complicated processes of quantum chemistry that may benefit from more formal examination.

Longuet-Higgins (1963), in a classic paper, argues 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 centre of mass.

As for IDP (Wallace, 2011b, c), we assume it possible to extend nonrigid molecule group theory to whip-like fronds anchored at one end via 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, 2011c and cited references). These are taken as parameterized by an index of 'frond length' L which might simply be the number of exposed monosaccharides. 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 Balasubraminian 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 glycan kelp frond. 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 is very precisely the 'topological catalysis' that decreases the 'friction' due to σ^2 in the self-lubrication argument, reexpressed in the language of nonrigid molecular symmetry.

Glycan 'sequence flexibility', i.e., the way in which intrinsic glycan functions can be carried out by an ensemble of possible glycan structures, a continuum, mirrors, somewhat, the dynamic matching of the fuzzy-key-and-lock mechanisms suggested by Tompa and Fuxreiter (2008). Taking the approach of Wallace (2012b), this can be addressed by supposing that the duality between a subgroup/subgroupoid of the glycan frond and the binding ligand site can be expressed as

$$\mathcal{B}_\alpha = C_\beta \mathcal{D}_\gamma \tag{7}$$

where $\mathcal{B}_\alpha$ is a subgroup/groupoid (or set of them) of the glycan nonrigid symmetry group or groupoid, $\mathcal{D}_\gamma$ a similar structure of the binding ligand, and C_β is an appropriate inversion operation or set of them that represents static or dynamic matching of the fuzzy key to the fuzzy lock. The fuzzyness, however, now extends to sequence replacement as well as variations in position, a far more complicated matter that may require an explicit extension of theory using groupoid methods and groupoid representations, as has been the case for stereochemistry (again, Wallace, 2011c and references). Groupoids are, in a sense, local symmetry structures that characterize the partial symmetries of finite tilings, quasicrystals, and the like.

4 Extending the model

An essential outcome of this approach is that the glycan symmetries, and their associated dynamics, should be highly punctuated in the parameter L that indexes final glycan topological form, and this effect should be observable.

For large L we apply a statistical mechanics analog for which we can use Landau's spontaneous symmetry breaking/lifting approach via a Morse theory argument, following Wallace (2012b). Typically, very many Morse functions are possible under a given circumstance, and we construct what is perhaps the simplest using group representations.

Taking an appropriate group representation in a particular matrix algebra, we can construct a 'pseudo probability' $\mathcal{P}$ for nonrigid group element ω as

$$\mathcal{P}[\omega] = \frac{\exp[-|\chi_\omega|/\kappa L]}{\sum_\nu \exp[-|\chi_\nu|/\kappa L]}, \tag{8}$$

where χ_ϕ is the character of the group element ϕ in that representation, i.e., the trace of the matrix assigned to ϕ, 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

$$\exp[-F/\kappa L] = \sum_\nu \exp[-|\chi_\nu|/\kappa L] \tag{9}$$

is a Morse Function in the temperature-analog L to which we can apply Landau's spontaneous symmetry breaking arguments (Pettini, 2007; Landau and Lifshitz, 2007), leading to the expectation of empirically-observable highly punctuated structure and reaction dynamics in the index L that are the analog to phase transitions in 'simple' physical systems: Landau's central insight was that, for many physical phenomena, raising the temperature would make accessible higher energy states of the system Hamiltonian, the quantum mechanical energy operator, and that the inherent symmetry changes would necessarily be punctuated. Here we focus directly on a Morse Function constructed from those symmetries, and rely on the robustness of the underlying mathematics to carry through, basically an empirical question.

5 Discussion and conclusions

We have extended the 'snake' model of intrinsically disordered proteins (Wallace, 2011a), as supplemented by a nonrigid molecule symmetry approach (Wallace, 2012b), to the flexible glycan/lectin system that coats the surface of the cell and provides an exceedingly rich channel for the transmission of biological information. A quantum chemistry context is implicit in the application of nonrigid molecule theory, although local tiling-matching may be characterized more by groupoid than group symmetries, as is the case for quasicrystals and groupoid stereochemistry (Wallace, 2011c and references therein).

The essential empirical implication is that sufficient search should reveal punctuated spectral transitions in glycan/lectin interaction that can be indexed by an appropriate measure of glycan 'length' – i.e., topological complexity – analogous to a generalized temperature. These spectral structures are in addition to the 'developmental' spectra predicted by Wallace (2012a) to characterize the glycan surface fronds themselves.

6 References

Anfinsen, C., 1973, Principles that govern the folding of protein chains, Science, 181:223-230.

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

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

Cover, T., H. 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.

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

Gabius, H., S. Andre, H. Kaltner, H. Siebert, 2002, The sugar code: functional lectinomics, Biochimica Biophysica Acta, 1572:165-177.

Gruebele, M., 2005, Downhill protein folding: evolution meets physics, Comptes Rendus Biologies, 328:701-712.

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

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

Longuet-Higgins, H., 1963, The symmetry groups of nonrigid molecules, Molecular Physics, 6:445-460.

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

Tlusty, T., 2007, A model for the emergence of the genetic code as a transition in a noisy information channel, Journal of Theoretical Biology, 249:331-342.

Tompa, P., M. Fuxreiter, 2008, Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions, Trends in Biochemical Science, 33:1-8.

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

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

Wallace, R., 2011b, Multifunction moonlighting and intrinsically disordered proteins: Information catalysis, nonrigid molecule symmetries, and the 'logic gate' spectrum, Comptes Rendus Chimie, 14:1117-1121.

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

Wallace, R., 2012a, Extending Tlusty's rate distortion index theorem method to the glycome: Do even 'low level' biochemical phenomena require sophisticated cognitive paradigms? In press, BioSystems, doi:10.1016/j.biosystems.2011.11.005.

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