

On biological homochirality

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

August 12, 2009

Abstract

Generalizing Landau's spontaneous symmetry breaking arguments using the standard groupoid approach to stereochemistry allows reconsideration of the origin of biological homochirality. On Earth, limited metabolic free energy density may have served as a low temperature-analog to 'freeze' the system in the lowest energy state, the set of simplest homochiral transitive groupoids representing reproductive chemistries. These engaged in Darwinian competition until a single configuration survived. Subsequent path-dependent evolutionary process locked-in this initial condition. Astrobiological outcomes, in the presence of higher initial metabolic free energy densities, could well be considerably richer, perhaps of mixed chirality. One result would be a complicated distribution of biological chirality across a statistically large sample of extraterrestrial stereochemistry, in contrast with a recent prediction of a racemic average.

Key Words: biological homochirality, groupoid, information theory, spontaneous symmetry breaking, stereochemistry

1 Introduction

Amino acids and the backbone of DNA/RNA in living things on Earth are found in only one of the two possible mirror-image states available to them. Respectively, the L-forms of amino acids primarily serve as the building blocks of proteins, and D-sugars form the DNA/RNA backbone [1]. Attempts to replicate early conditions on Earth – Miller/Urey experiments – always produce 'racemic' mixtures having equal amounts of both possible amino acid symmetry forms. This fundamental conundrum was recently addressed by Gleiser et al. [2], who, in a computationally intensive study adapting Sandars' 'toy model' of polymerization [3], conclude that other planetary platforms in this solar system and elsewhere could have developed an opposite chiral bias to that of Earth. As a consequence, a statistically large sampling of extraterrestrial stereochemistry would be necessarily racemic on average.

Here we attempt a more direct treatment based on the homology between free energy density and information source uncertainty that was the basis of the analysis used in [4], and argue, via a statistical thermodynamic construction, that

available metabolic energy could well have been the principal determining environmental influence. As a consequence of groupoid symmetries associated with stereochemical structure, a statistically large sampling of extraterrestrial stereochemistries could well be far more complex than [2] proposes, i.e., not racemic on average.

The development is straightforward:

1. Reproducing molecular codes, in the largest sense, themselves constitute information sources that are Darwinian individuals, subject to variation, selection, and chance extinction.
2. Enantiomeric forms of molecules constitute equivalence classes that can be represented by groupoid, rather than group, symmetries, leading to a groupoid version of Landau's classic phenomenological model for phase transition and its extension via Pettini's 'topological hypothesis' [5]. The necessity of using groupoid methods in stereochemistry has long been recognized, and will not be reviewed here [6-13]. For a tutorial on groupoid methods see Brown [14] or Weinstein [15]. Wallace and Wallace [16] provide other details.
3. Groupoid symmetries and available metabolic free energy are, as a consequence of the Darwinian individuality of coding schemes, in the sense of Gould [17], contexts for, rather than determinants of, the resulting evolutionary processes, including punctuated equilibrium.

A strong current in contemporary theoretical biology [18-20] argues that, in modern organisms, genomic complexity fits within standard information theory as the information the genome of an organism contains about its environment, so that evolution on the molecular level is a collection of information transmission channels, subject to certain constraints defined by the asymptotic limit theorems of information theory. The organism's genes code for the information, a message, to be transmitted from progenitor to offspring, and are subject to noise from an imperfect reproduction process. Thus the information content, or complexity, of a genomic string by itself, without referring to the embedding environment, is a meaningless concept, and a change in environment leads to a change in complexity. The transmission of reproductive information is thus a contextual matter involving the operation of an information source that must interact with embedding ecosystem structures. Here we will focus on the role of available metabolic free energy as the main driving environmental factor.

Reproduction – biotic or prebiotic – is thus to be char-

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

acterized by an information source, whose source uncertainty has an important heuristic interpretation. Following Ash [21], we may regard a portion of text in a particular language as being produced by an information source. The probabilities $P[X_n = a_n | X_0 = a_0, \dots, X_{n-1} = a_{n-1}]$ may be estimated from the available data about the language; in this way we can estimate the uncertainty associated with the language. A large uncertainty means, by the Shannon-McMillan Theorem, a large number of ‘meaningful’ sequences. Thus given two languages with uncertainties H_1 and H_2 respectively, if $H_1 > H_2$, then in the absence of noise it is easier to communicate in the first language; more can be said in the same amount of time. On the other hand, it will be easier to reconstruct a scrambled portion of text in the second language, since fewer of the possible sequences of length n are meaningful.

Thus, depending on the degree of noise, either high or low reproductive source uncertainty can have selective advantage, a kind of stochastic resonance related to the mesoscale resonance arguments of [16].

2 Free Energy Density and Information Source Uncertainty

Information source uncertainty can be defined in several ways. Khinchin [22] describes the fundamental ‘E-property’ of a stationary, ergodic information source as the ability, in the limit of infinity long output, to classify strings into two sets:

1. a very large collection of gibberish which does not conform to underlying rules of grammar and syntax, in a large sense, and which has near-zero probability, and
2. a relatively small ‘meaningful’ set, in conformity with underlying structural rules, having very high probability.

The essential content of the Shannon-McMillan Theorem is that, if $N(n)$ is the number of ‘meaningful’ strings of length n , then the uncertainty of an information source X can be defined as

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

$$\lim_{n \rightarrow \infty} H(X_n | X_0, \dots, X_{n-1}) =$$

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

(1)

where $H(\dots|\dots)$ and $H(\dots)$ are conditional and joint Shannon uncertainties defined from the appropriate cross-sectional string probabilities.

The free energy density of a physical system having volume V and partition function $Z(K)$ derived from the system’s Hamiltonian at inverse temperature K is [23]

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

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

(2)

where $\hat{Z} = Z^{-1/K}$.

Feynman [24] concludes that the information contained in a message is simply the free energy needed to erase it. Thus, according to this argument, source uncertainty is homologous to free energy density as defined above.

Ash’s remark then has a corollary: If, for a biological system, $H_1 > H_2$, source 1 will require more metabolic free energy for ongoing maintenance than source 2.

3 The Basic Model

We begin by classifying the available molecules in our prebiotic soup by their underlying stereochemistries, and allow the reproductive systems to, for purposes of initial classification, reflect those stereochemical equivalence classes. Interactions between stereochemical equivalence classes can be used to classify higher order structures.

Equivalence classes define groupoids, by standard mechanisms [14, 15, 25-27]. The basic equivalence classes will define transitive groupoids, and higher order systems can be constructed by the union of transitive groupoids, having larger chemical alphabets that allow more complicated statements in the sense of Ash above.

Given an appropriately scaled, dimensionless, fixed, inverse available metabolic energy density K , we propose that the metabolic-energy-constrained probability of a reproductive information source representing stereochemical equivalence class D_i , H_{D_i} , will be given by the classic relation [23]

$$P[H_{D_i}] = \exp[-H_{D_i}K] / \left[\sum_j \exp[-H_{D_j}K] \right],$$

(3)

where the sum is over all possible elements of the largest available symmetry groupoid. By the arguments above, compound sources, formed by the union of (interaction of species

from) underlying transitive groupoids, being more complex, will all have higher free-energy-density-equivalents than those of the base (transitive) groupoids.

Let

$$Z_D \equiv \sum_j \exp[-H_{D_j}K].$$

(4)

We now define the *Groupoid free energy* of the system, F_D , at inverse normalized metabolic energy density K , as

$$F_D[K] \equiv -\frac{1}{K} \log[Z_D[K]].$$

(5)

We have expressed the probability of a reproductive information source in terms of its relation to a fixed, scaled, available (inverse) metabolic free energy density seen as a kind of equivalent (inverse) system temperature. This gives a statistical thermodynamic path leading to definition of a ‘higher’ free energy construct – $F_D[K]$ – to which we now apply Landau’s fundamental heuristic phase transition argument [5, 23, 28].

The essence of Landau’s insight was that second order phase transitions were usually in the context of a significant symmetry change in the physical states of a system, with one phase being far more symmetric than the other. A symmetry is lost in the transition, a phenomenon called spontaneous symmetry breaking. The greatest possible set of symmetries in a physical system is that of the Hamiltonian describing its energy states. Usually states accessible at lower temperatures will lack the symmetries available at higher temperatures, so that the lower temperature phase is less symmetric: The randomization of higher temperatures – in this case higher available metabolic free energy densities – ensures that higher symmetry/energy states – mixed transitive groupoid structures – will then be accessible to the system. Absent high metabolic free energy densities, however, only the simplest transitive groupoid structures can be manifest, i.e., those associated with the simplest stereochemistries. A full treatment from this perspective requires invocation of groupoid representations [29, 30].

An extended version of Pettini’s [5] Morse-Theory-based topological hypothesis can now be invoked, i.e., that changes in underlying groupoid structure are a necessary (but not sufficient) consequence of phase changes in $F_D[K]$. Necessity, but not sufficiency, allows mixed symmetries, e.g., L-forms of amino acids working in concert with the D-sugar DNA/RNA backbone.

4 Discussion and Conclusions

What, in one sense, compromises this analysis, but in a more fundamental way completes it, is that the reproductive chemical strategies represented by the H_{D_j} are not merely passive actors. Quite the contrary, they are full-scale Darwinian individuals in the sense of [17], and thus subject to variation, selection, and chance extirpation. Thus, given sufficient initial metabolic energy density, there is no inherent reason why higher order, non-transitive, groupoid reproductive chemical systems – of mixed chirality – might not prevail, particularly in view of the Ash quotation above. That is, one can ‘say’ more in a shorter time using a richer reproductive language, and this might well have selective value. Thus we may, if this model is correct, expect to observe some surprising astrobiological reproductive stereochemistries, in contrast to the simple ‘racemic’ conclusion of [2].

The corollary to this argument is that initial preaerobic metabolic free energy density on Earth may just not have been sufficient to activate non-homochiral reproductive chemistries, and that the two possible amino acid systems, L, D , engaged in a competition through which one prevailed. Subsequent path-dependent evolutionary lock-in produced the ultimate result.

Again, groupoid symmetries and available metabolic free energy are, as a consequence of the Darwinian individuality of reproductive coding schemes, contexts for, rather than determinants of, evolutionary process, including punctuated equilibrium. They are the banks between which the prebiotic evolutionary glacier flowed – sometimes slowly, and sometimes in sudden advance.

5 References

1. Fitz, D., H. Reiner, K. Plakensteiner, and B. Rode, 2007, Possible origins of biohomochirality, *Current Chemical Biology*, 1:41-52.
2. Gleiser, M., J. Thorarinson, and S. Walker, 2008, Punctuated chirality, *Origins of Life and Evolution of Biospheres*, DOI 10.1007/s11084-008-9147-0.
3. Sandars, P., 2003, A toy model for the generation of homochirality during polymerization, *Origins of Life and Evolution of Biospheres*, 33:575-587.
4. Wallace, R., and R.G. Wallace, 2008, On the spectrum of prebiotic chemical systems: an information-theoretic treatment of Eigen’s paradox, *Origins of Life and Evolution of Biospheres*, 38:419-455.
5. Pettini, M., 2007, *Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics*, Springer, New York.
6. Dornberger-Schiff, K., and H. Grell-Niemann, 1961, On the theory of order-disorder structures, *Acta Crystallographica*, 14:167-177.
7. Klemperer, W., 1973, The steric courses of chemical reaction, *Journal of the American Chemical Society*, 95:380-396.

8. Nourse, J., 1975, An algebraic description of stereochemical correspondence, *Proceedings of the National Academy of Sciences*, 72:2385-2388.
9. Sadanaga, R., and K. Ohsumi, 1979, Basic theorems of vector symmetry in crystallography, *Acta Crystallographica A*, A35:115-122.
10. Fichtner, K., 1986, Non-space-group symmetry in crystallography, *Computers and Mathematics with Applications*, B, 12:751-762.
11. Yamamoto, A., and K. Ishihara, 1988, Penrose patterns and related structures. II. Decagonal quasicrystals, *Acta Crystallographica A*, A44:707-714.
12. Cayron, C., 2006, Groupoid of orientational variants, *Acta Crystallographica A*, A62, 21-40.
13. Cayron, C., 2007, Multiple twinning in cubic crystals: geometric/algebraic study and its application for the identification of the $\Sigma 3^n$ grain boundaries, *Acta Crystallographica A*, A63:11-29.
14. Brown, R., 1987, From groups to groupoids: a brief survey, *Bulletin of the London Mathematical Society*, 19:113-134.
15. Weinstein, A., 1996, Groupoids: unifying internal and external symmetry, *Notices of the American Mathematical Association*, 43:744-752.
16. Wallace, R., and D. Wallace, 2008, Punctuated equilibrium in statistical models of generalized coevolutionary resilience: how sudden ecosystem transitions can entrain both phenotype expression and Darwinian selection, *Transactions on Computational Systems Biology IX*, LNBI 5121:23-85.
17. Gould, S., 2002, *The Structure of Evolutionary Theory*, Harvard University Press, Cambridge, MA.
18. Ofria, C., C. Adami, and T. Collier, 2003, Selective pressures on genomes in molecular evolution, *Journal of Theoretical Biology*, 222:477-483.
19. Adami, C., and N. Cerf, 2000, Physical complexity of symbolic sequences, *Physica D*, 137:62-69.
20. Adami, C., C. Ofria, and T. Collier, 2000, Evolution of biological complexity, *Proceedings of the National Academy of Sciences*, 97:4463-4468.
21. Ash R., 1990, *Information Theory*, Dover Publications, New York.
22. Khinchin, A., 1957, *Mathematical Foundations of Information Theory*, Dover, New York.
23. Landau, L., and E. Lifshitz, 2007, *Statistical Physics, 3rd Edition*, Part I, Elsevier, New York.
24. Feynman, R., 2000, *Lectures on Computation*, Westview Press, New York.
25. Cannas Da Silva, A., and A. Weinstein, 1999, *Geometric Models for Noncommutative Algebras*, American Mathematical Society, RI.
26. Wallace, R., 2007, Culture and inattentional blindness, *Journal of Theoretical Biology*, 245:378-390.
27. Golubitsky, M., and I. Stewart, 2006, Nonlinear dynamics and networks: the groupoid formalism, *Bulletin of the American Mathematical Society*, 43:305-364.
28. Skierski, M., A. Grundland, and J. Tuszynski, 1989, Analysis of the three-dimensional time-dependent Landau-Ginzburg equation and its solutions, *Journal of Physics A (Math. Gen.)*, 22:3789-3808.
29. Buneci, M., 2003, *Representare de Grupoizi*, Editura Mirton, Timisoara.
30. Bos, R., 2007, Continuous representations of groupoids, arXiv:math/0612639.