CHIRALITY 21:354–358 (2009)

A New Definition of Life

JAMES D. CARROLL*

ABSTRACT Chirality is often glossed over in theoretical or experimental discussions concerning the origin of life, but the ubiquity of homochiral building blocks in known biological systems demands explanation. Information theory can provide a quantitative framework for understanding the role of chirality in biology. Here I show how conclusions derived from information theory, in particular the concept of equivocation, can explain not only why chiral building blocks are necessary in living systems but also why a homochiral set of building blocks is necessary. These results lead to a new definition of life, and to the conclusion that the simplest form of life exists in the form of self-amplifying, autocatalytic reactions such as the Soai reaction. *Chirality 21:354–358, 2009.*

KEY WORDS: homochirality; information theory; self-replication; Soai reaction; origin of life

INTRODUCTION

Answers to questions concerning the origin of life are complicated by an absence of fundamental postulates. Without any fundamental postulates for guidance, the kind and number of theories that are possible to conceive are probably limitless. The specialization of scientists in their respective fields is also a difficulty because the perspective of their specialization can limit their ideas. For example, to biochemists, the so-called RNA world[1] that has been postulated as the precursor to the known forms of life may appear to them to be a thankfully simpler version of the extant biology, but to a small molecule synthetic organic chemist, it remains hopelessly complex with respect to origins. How does a single molecule of ribose, with its complex stereochemistry, form, much less the large polymers of biology? Likewise, some theorists have postulated small-molecule chemical schemes without any inclusion or explanation of stereochemistry.[2] However, if we consider the molecular origin of life from the perspective of information theory, the issues become clearer.

Information theory has been applied to biological problems in the past; there is an entire book on the subject by Yockey,[3] but only part of the information content of biomolecules was considered and his major conclusion is essentially negative. He concludes that the origin of the genetic code is "unknowable"[4] and (the italics are his):

"No code exists to send information from protein sequences to mRNA or DNA .... Therefore it is *impossible* that the origin of life was proteins first ...."[5]

This conclusion was reached by considering DNA as a string of letters only (... ATTGCAAGC ...) and likewise by considering proteins as strings of identifiers (... DYRFQ ...), and while his quantification of the information content of proteins considered as such is certainly correct, that particular conclusion is probably wrong because he failed to consider entirely the information content of the molecular structures themselves. For example, DNA has a polymeric backbone of deoxyribose. This particular sugar may consist of eight possible stereoisomers, and so the information necessary to construct this particular isomer (and its ribose precursor, which may exist as 16 possible stereoisomers) must exist in addition to the information which encodes the sequences of amino acids which form the enzymes that assemble ribose in the cell. In addition, the information content of proteins which exists apart from their sequences of amino acids, for example, the information content of their conformations, was not considered. Researchers have already begun to quantify the information content of conformations,[6] and given this and other information that is contained in the molecular structure of proteins, it certainly is possible that proteins came first. I will consider here another form of information present in proteins and other biomolecules: that which is implied by their chirality. First, a brief digression concerning information theory is necessary.

INFORMATION THEORY

The simple example of the coin flip is often used to introduce quantitative information theory. A coin flip of an unbiased coin can have two results, heads or tails. How many binary bits are required to convey the result of the flip? Shannon[7] found that the amount of information (designated by "*H*") is given by eq. 1:

*Correspondence to: James D. Carroll, P.O. Box 764, Newburyport, MA 01950, USA. E-mail: jcarroll99@aol.com
Received for publication 17 February 2008; Accepted 15 April 2008
DOI: 10.1002/chir.20590
Published online 20 June 2008 in Wiley InterScience (www.interscience.wiley.com).

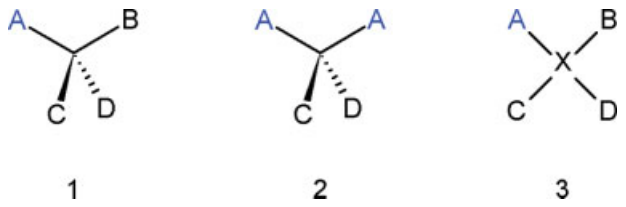

Fig. 1. Tetrahedral and Square Planar Substitutions.

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i \quad (1)$$

where the p_i are the probabilities of each event (or symbol) that occurs. In the case of the coin flip, H is given by eq. 2:

$$-\left[\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right] = 1 \quad (2)$$

Thus, 1 bit, either a 0 or a 1, is sufficient to convey the information. Note that H is only a quantity of information, and says nothing about the knowledge contained in the information, for example, if heads wins the last bottle of beer, or whether a 0 should represent heads or tails, or even if a coin flip is involved at all. To use eq. 1 in a chemical setting, we could, for example, follow the method of Sullivan (Ref. 6) who calculates the information content of molecular conformations by assigning a probability to each possible conformation and summing as in eq. 1.

CHIRALITY AND INFORMATION THEORY

A chiral carbon molecule has four different substituents attached to it, as in Figure 1. Upon first consideration, a method of calculating the amount of information contained in **1** might seem to be analogous to the calculation of the information given by a four-sided die. The information content of such a die is given by eq. 3:

$$H = -\sum_{1}^{4} \frac{1}{4}\log_2\frac{1}{4} = 2 \quad (3)$$

And likewise the information content of a carbon molecule **2** with C_s symmetry (such as glycine) would be given by eq. 4 (as for a die with 1,1,2,3 on its faces):

$$H = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{4}\log_2\frac{1}{4} = 1.5 \quad (4)$$

However, a hypothetical square planar molecule such as **3** also has 4 different substituents and contains the same amount of substituent information as **1**. Note that while an individual amino acid has the information content given by eq. 3, in the context of a peptide, three of the groups contribute no information. This is because three of the groups of each amino acid are the same, and therefore the substituent information content of the peptide resides solely in the side chain of each amino acid. An analogy could be made to our own language if we were to precede each letter in a word by the same three symbols, for example, rather than "cat", we wrote "@#$c@#$a@#$t"; the information resides solely in the last symbol of each four and the rest can be ignored. If the frequencies of each amino acid were equal, the information content would be $\log_2(1/20)$ for each residue. This substituent information is the information discussed by Yockey in Ref. 3. What I am considering and what we need to be concerned with in analyzing the existence of chiral building blocks in biology, is the spatial information contained in structures **1–3**.

Spatial information would be impossible to quantify (as any other quantity other than infinity) because in any coordinate system the coordinates may be subdivided infinitely, and thus the information content of a free object would be infinite. However, molecules such as **1–3** have constraints of bond lengths and geometries such that a finite number of spatial coordinates specifies the location of the substituents relative to an origin placed at the center of the point group. If we place the origin at the chiral center of **1**, for example, and if the bond lengths from the central carbon are assumed to be equal (equivalent to requiring **r** to be constant in a spherical polar coordinate system), then two angles, θ and ϕ, suffice to locate the first atom of each substituent. Two pairs of such angles exist, corresponding to each enantiomer, which we may label as $\theta_A\theta_B\theta_C\theta_D$, $\theta_A\theta_B\theta_D\theta_C$ and $\phi_A\phi_B\phi_C\phi_D$, $\phi_A\phi_B\phi_D\phi_C$, where, for example, θ_A represents the angle θ that the group labeled A occupies in Figure 1, and the switching of the coordinates of the last two groups occurs because exchanging any two groups of structure **1** creates its enantiomer. Thus, the spatial information content of chiral carbon molecules with a single chiral center is given by eq. 5:

$$H = -\sum_{1}^{2} \frac{1}{2}\log_2\frac{1}{2} - \sum_{1}^{2} \frac{1}{2}\log_2\frac{1}{2} = 2 \quad (5)$$

For carbon molecules with more than one chiral center, the number of unique sets of coordinates increases as powers of 2, and the information content therefore increases as multiples of 2 (excluding meso forms). For example, a carbon molecule with 4 chiral centers has 2^4 or 16 possible unique sets of coordinates, and therefore $\log_2(2^4)$ or 4 bits per coordinate or 8 total bits of spatial information. In other words, the total number of bits is the number of chiral centers times 2.

Structure **2** contains no spatial information at all, since every permutation of substituents is superimposable and thus there is only one set of the θs and ϕs, and $1*\log(1) = 0$. The square planar molecule **3** consists of three distinct isomers, corresponding to those with AB, AC, and AD on the diagonal, respectively, and therefore there is spatial information concerning one dimension only equal to that given by eq. 6:

$$H = -\sum_{1}^{3} \frac{1}{3}\log_2\frac{1}{3} = 1.58 \quad (6)$$

For planar molecules such as **3**, even if there were many more substituents in the plane, the number of bits of infor-

mation would increase, but the information concerning the third spatial dimension would always be zero.

Why is this important? Because all biological life consists of three-dimensional objects, and if the building blocks of life contained no three-dimensional information, far more molecules would be required to specify a particular three-dimensional object (such as the active site of an enzyme, and, ultimately, the overall three-dimensional shape of the organism). For example, the square planar molecule 3 can provide no unambiguous spatial information out of the plane of the molecule, so another planar molecule would have to be oriented out of the plane of the first one in order to specify a position out of the plane unambiguously. In other words, for life which consists of carbon molecules, chiral molecules are the most efficient building blocks for storing spatial information.

An example of a spatially inefficient self-replicating molecular system may be self-replicating micelles.[8] Self-replicating micelles consisting of up to 380 *n*-octanoate molecules and having a molecular weight of up to ~54 kDa were reported. Despite the size and complexity of such entities, they are capable of forming only simple geometric shapes such as spheres or cylinders because there is so little spatial information in each molecule of octanoate. All of the carbon atoms in the octanoate molecule have symmetry, and therefore the two types of functional groups connected together (the polar "head" and the nonpolar "tail") within the the molecule provide one-dimensional information that signifies a head-to-tail line, and the nonpolar tail must consist of multiple atoms to do so. In contrast, small enzymes such as 4-oxalocrotonate tautomerase,[9] which consists of 76 amino acid residues with a molecular weight of only ~8.5 kDa, form complex three-dimensional shapes that have significant biological functions.

WHY HOMOCHIRALITY?

We now need to consider why a homochiral set of building blocks appears to be required for biological life, as opposed to a racemic set of chiral molecules or a set of some other %ee. This question can be answered by drawing upon another concept from information theory, that of equivocation.[10] This concept was originally applied to so-called noisy information channels in order to calculate how much information may be transmitted over a channel in cases where the reception of transmitted information was unreliable. For example, if we return to the example of the coin toss, we may have a channel as in Figure 2, where heads or tails (0 or 1) is transmitted with a 50% chance for each case that the opposite bit may be received.

If by $H(x)$ we mean the quantity of information transmitted, and by $H(y)$ the quantity received, then the equivocation, $H_y(x)$, is given by eq. 7:

$$H_y(x) = -\sum_{y=1}^{m}\sum_{x=1}^{m} p(y)p_y(x)\log_2 p_y(x) \qquad (7)$$

where $p(y)$ is the probability that y is received and $p_y(x)$ is the uncertainty that x was sent when y is received. In the case of the coin toss described above, we have:

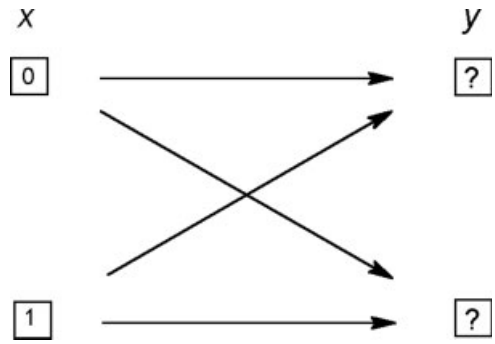

Fig. 2. Symbolic representation of a communications channel with equivocation.

$$H_y(x) = -\frac{1}{2}\left(\frac{1}{2}\log_2\frac{1}{2}+\frac{1}{2}\log_2\frac{1}{2}\right) - \frac{1}{2}\left(\frac{1}{2}\log_2\frac{1}{2}+\frac{1}{2}\log_2\frac{1}{2}\right) = 1 \qquad (8)$$

So the information capacity of the channel in Figure 2 is zero because $H(y) = H(x) - H_y(x) = 1 - 1 = 0$.

Consider now a peptide made up of one amino acid such as alanine. If the alanine is enantiomerically pure, another molecule "reading" the spatial information of the side chain (in this case, a methyl group) unambiguously "receives" the information that a methyl group in the chain is on one side of the amide bond chain. But suppose we use a 90:10 mixture (80% ee) of (+) and (–) alanine. Each amino acid in the chain contributes 1 bit of spatial information in the θ dimension (and likewise for the ϕ dimension), but the equivocation for each amino acid is:

$$\begin{aligned} H_y(x) = &-0.90(0.90\log_2 0.90 + 0.10\log_2 0.10) \\ &- 0.10(0.90\log_2 0.90 + 0.10\log_2 0.10) \\ &= 0.53 \end{aligned} \qquad (9)$$

So the information in the θ dimension transmitted per amino acid is reduced to 0.47 bits, and for the case of a racemic mixture, to zero. Thus, a homochiral set of building blocks is the most efficient way to encode spatial information, and as before in the case of achiral building blocks, any reduction in the amount of spatial information supplied by the building blocks would have to make up by more atoms.

One objection to the above arguments could be: "What if an enzyme were selective only for a particular diastereomer formed from a heterochiral pool of amino acids?" But information theory assumes that there is one "receiver" which interprets the received signal unambiguously. This is referred to sometimes as the "ideal observer."[11] Returning again to the example of the coin flip, if we have a noiseless channel but have two receivers, one that interprets a 1 as a 1 and another that interprets a 1 as a 0, we have in effect created a noisy channel, the output of which is identical to that shown in Figure 2, and this particular case is equivalent to reception of a 1 as a 1 and 0 as 0 with 50% fidelity, so the equivocation would be 1 and the amount of useful information transmitted over

such a channel would again be 0. A heterochiral pool of enantiomers would not consist of one enzyme that interprets the spatial information of a particular diastereomer and not others, but of multiple analogous enzymes which would interact in different ways with a particular diastereomeric substrate.

A more pertinent objection can be raised by asking "What if there were a pool of chiral building blocks, each enantiomerically pure, but not homochiral?" For example, suppose of the 19 chiral amino acids, 10 were (L) and 9 were (D). In that case we would not have multiple diastereomeric forms for a given chain of amino acids, but one diastereomer only for each chain. This question goes to the root of the issue of what it means to say that chiral molecules "interact" or exchange information. What precisely is this information? The answer is that a given chiral building block defines the coordinate system. It determines "left"/"right," "up"/"down," and "in"/"out", or the positive and negative directions of any three-dimensional coordinate system that is preferred. For example, suppose we wish to construct an arbitrary three-dimensional shape as shown in Figure 3 (the solid lines are in the plane of the page or parallel to it, the wedges have their usual meaning):

In the coordinate system of the drawing, molecule 2 has been attached "right" from molecule 1, and 3 is attached to the "bottom" of molecule 2, 4 "out" relative to 3, etc. But how is the "constructor" (presumably an enzyme made of chiral building blocks) to define precisely what those directions mean relative to the initial building block? It does so by reference to its own set of analogously substituted, homochiral building blocks. In the case of amino acids, the relative positions of the side chain, hydrogen, carboxyl group and amino group define in a consistent manner where is "where," A heterochiral mix of such amino acids, either of low % ee or 100% ee but of mixed handedness, would create equivocation of the information concerning location. It is essential therefore to separate the spatial information contained in the amino acids from the information provided by the functional groups of the side chains. As far as spatial information is concerned, all of the amino acids of the homochiral set are identical entities. The implication of all of this is that homochirality can be thought of as the 21st "functional group" of the amino acid building blocks. It may be difficult at first to imagine what the equivalent of heterochirality would be for functional groups, but consider a functional group such as the $-CH_2OH$ group of serine. Its "enantiomer" would have to be another CH_2OH group, alike in every way, but with the property that it repels hydrogen rather than hydrogen bonds to it!

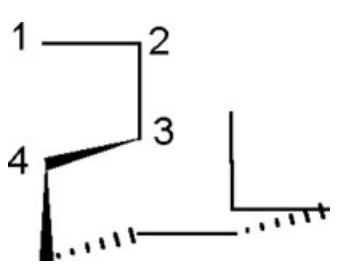

Fig. 3. An arbitrary three-dimensional object.

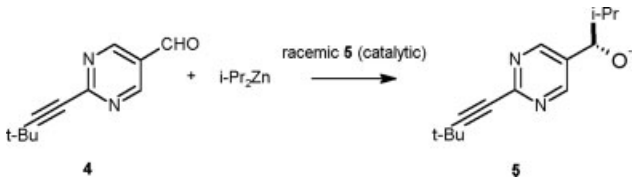

Fig. 4. The Soai reaction.

THE SIMPLEST FORM OF CARBON-BASED LIFE

If the foregoing arguments are correct, then before any efficient (in terms of molecular size) life forms may arise, a pool of homochiral building blocks must be available. The chicken-egg dilemma is resolved by assuming that the origin of homochirality and the origin of life were *the same event*. Further, before a homochiral set of building blocks arose, one homochiral building block was made available. A reaction that could serve as a source of a homochiral building block is exemplified by the Soai reaction.[12] This reaction produces, with asymmetric amplification, pyrimidyl alcohols as in Figure 4 (the acidic workup is left out to save space so **5** is depicted as the alkoxide).

The alcohol **5** is therefore replicating itself stereoselectively. Blackmond has discussed the relevance of the Soai reaction to the origin of biological homochirality.[13] Kinetic studies[14] indicate that the catalytic entity in the reaction may be tetrameric, involving two molecules of homochiral reduced substrate and two molecules of mono-isopropyl zinc (heterochiral species react more slowly or not at all, and this is the source of the asymmetric amplification). The inclusion of two homochiral alkoxide molecules in the catalyst thus fulfills one of the requirements for self-replicating automata pointed out by Von Neumann[15]:

"When an automaton performs certain operations, they must be expected to be of a lower degree of complication than the automaton itself. In particular, if an automaton has the ability to construct another one, there must be a decrease in complication as we go from the parent to the construct."

In the case of the Soai reaction, the extra "complication" is, in part, the information content of the two homochiral ligands, which results in the production of one molecule of product of the same enantiomer. Further, the reaction kinetics appears to be analogous to the Michaelis-Menton kinetics of enzymes.[16] This is only a type example, because the zinc reagent is unstable in water, but analogies to or contradictions with the physical construction of cellular life are less important than the more general analogy represented by the transmission of three-dimensional information from one molecule to another.

CONCLUSIONS

Numerous definitions of life have been proposed in the past,[17] but the conclusions of this article are clear. First, a

fundamental unit of information of life may not be the DNA codon, but rather the bits of information that represent the unambiguous spatial information supplied by the chirality of the molecules involved, in addition to the information supplied by the functional groups of protein side chains. Second, a new definition of life suggests itself: *life is that which self-reproduces a homochiral environment*. Note that this definition does not require that every building block be chiral, only that those molecules which are chiral be homochiral. Achiral glycine has a purpose as a constituent of proteins because of its low steric bulk. This allows it to be a turning point in proteins which have symmetric or other unusual shape requirements, and therefore proteins which have a relatively high glycine content have atypical shapes for enzymes, such as the antifreeze proteins of snow fleas, which form a symmetric grouping of helices which bind to certain planes of ice crystals,[18] or the antibiotic protein microcin B17, in which glycine residues form what are essentially strings which link together the biologically active oxazole and thiazole rings that are formed post-translationally from adjacent serine and cysteine residues.[19] Finally, the implication for theorists of the origin of life is equally clear: theoretical or experimental work which ignores homochirality and its origin is at best incomplete.

LITERATURE CITED

1. Muller UF. Re-Creating an RNA World. Cell Mol Life Sci 2006;63:1278–1293.
2. Wachtershauser G. Life as we don't know it. Science 2000;289:1307–1308.
3. Yockey HP. Information theory, evolution, and the origin of life. New York: Cambridge University Press; 2005.
4. Yockey HP. Information theory, evolution, and the orgin of life. New York: Cambridge University Press; 2005, p 94.
5. Yockey HP. Information theory, evolution, and the orgin of life. New York: Cambridge University Press; 2005, p 147.
6. Sullivan DC, Tiba A, Voelz VA, Kuntz JD. Information content of molecular structures. Biophys J 2003;85:174–190.
7. Shannon CE. A mathematical theory of communication. Bell System Tech J 1948;27:379–423.
8. Bachmann PA, Walde P, Luisi PL, Lang J. Self-replicating micelles: aqueous micelles and enzymatically driven reactions in reverse micelles. J Am Chem Soc 1991;113:8204–8209.
9. Almrud J, Kern A, Wang S, Gerwinski R, Johnson W, Murzin A. The crystal structure of YdcE, a 4-oxalocrotonate tautomerase homologue from *Escherichia coli*, confirms the structural basis for oligomer diversity. Biochemistry 2002;41:12010–12024.
10. Pierce JR. An introduction to information theory: symbols, signals and noise. New York: Dover Publications; 1980. p 145–159.
11. Ash RB. Information theory. New York: Dover Publications; 1990. p 60–63.
12. Soai K, Shibata T, Morioka H, Choji K. Asymmetric autocatalysis and amplification of enantiomeric excess of a chiral molecule. Nature 1995;378:767–768.
13. Blackmond DG. Asymmetric autocatalysis and its implications for the origin of homochirality. Proc Natl Acad Sci USA 2004;101:5723–5736.
14. Blackmond DG. Mechanistic study of the Soai autocatalytic reaction informed by kinetic analysis. Tetrahedron: Asymmetry 2006;17:584–589.
15. Von Neumann J.Collected Works, Volume 5. New York: Pergamon Press; 1961. p 312.
16. Micskei K, Pota G, Caglioti L, Palyi G. Empirical description of chiral autocatalysis. J Phys Chem A 2006;110:5982–5984.
17. Palyi G, Zucchi G, Caglioti L, editors. Fundamentals of life (Section 1: Definitions of life). Elsevier Life Sciences: Paris; 2002. p 15–202.
18. Lin F-H, Graham LA, Campbell RL, Davies PL. Structural modeling of snow flea antifreeze protein. Biophys J 2007;92:1717–1723.
19. Bayer A, Freund S, Jung G. Post-translational heterocyclic backbone modifications in the 43-peptide antibiotic microcin B17. Structure elucidation and NMR study of a 13C,18N-labelled gyrase inhibitor. Eur J Biochem 1995;234:414–426.