Comment on ‘A colorful origin for the genetic code: Information theory, statistical mechanics and the emergence of molecular codes’ by T. Tlusty

Rodrick Wallace

Division of Epidemiology, The New York State Psychiatric Institute, Box 47, 1051 Riverside Dr., New York, NY 10032, United States

Received 26 July 2010; accepted 28 July 2010
Available online 3 August 2010
Communicated by E. Shakhnovich

It is difficult to exaggerate the importance of Tlusty’s application of rate distortion and topological arguments to the genetic code [1–5], not only for insights regarding the code itself, but for possible applications to a broad class of biological phenomena associated with information transmission. Although not emphasized by the review, a rate distortion approach to the genetic code can be rigorously restated in more traditional information theory terms and generalized to nonequilibrium dynamics constrained by the availability of environmental metabolic free energy [6]. But the genetic code is only the first of a large nested set of biological information processes, characterized, in its second step, by protein production also constrained by rate distortion dynamics and metabolic free energy [7]. What may be of particular interest, however, is an application of Tlusty’s topological arguments to the ‘protein folding code’. As Kamtekar et al. [8] point out, experimental studies of natural proteins show how their structures are remarkably tolerant to amino acid substitution, but that tolerance is limited by a need to maintain the hydrophobicity of interior side chains. Thus, while the information needed to encode a particular protein fold is highly degenerate, this degeneracy is constrained by a requirement to control the locations of polar and nonpolar residues. This is the precise protein folding analog to Tlusty’s error network analysis of Section 3, and the coloring arguments should thus apply, in some measure, to protein folding as well. Normal irregular protein symmetries were first classified by Levitt and Chothia [9], following a visual study of polypeptide chain topologies in a limited dataset of globular proteins. Four classes were observed; all α-helices; all β-sheets; α/β; and α + β, with the obvious interpretations. While this scheme strongly dominates observed irregular protein forms, heroic work by Chou and Cai [10] on a massive dataset recognizes three more symmetry equivalence classes; μ (multi-domain); σ (small protein); and ρ (peptide). Generalizing Tlusty’s Table 1 according to the genus γ of the underlying graph, that is, the number of holes, we obtain

DOI of original article: 10.1016/j.plrev.2010.06.002.
E-mail address: wallace@pi.cpmc.columbia.edu.
Thus we infer, from [9,10], an underlying ‘protein folding code error network’ that is, essentially, a large connected ‘sphere’ having one or more very small attachment handles, in the Morse Theory sense [11], a matter opening up other analytic approaches.

It should be possible to follow Tlusty’s methods much further up the nested hierarchy of biological information transmission, and this indeed represents a scientific open season. Good hunting!

References