Journal of Biomolecular Structure & Dynamics, ISSN 0739-1102
Volume 29, Issue Number 4, February 2012

Comment

Defining Life: An Exercise in Semantics or A Route to Biological Insights?

Eugene V. Koonin[1]

[1]National Center for Biotechnology Information, National Library of Medicine, National Institute of Health, Bethesda, MD 20894

http://www.jbsdonline.com

Asking 'What is X?' questions is a natural human inclination, and because all scientists that have so far published are undoubtedly human, they discuss such issues often enough. 'What is Life?' almost by definition is the king of such questions as far as Biology is concerned (1). Indeed, in the article which I address in this commentary (2), Edward Trifonov cites a recent special issue of the journal *Origins of Life and Evolution of Biospheres* that consists of 16 articles fully dedicated to different aspects of defining Life and the features of the resulting definitions (3). Certainly, this is evidence of considerable attention to the subject. Yet, in itself the question 'What is Life?' hardly can be considered scientific. Falsification is impossible: whenever one finds an apparent counterexample, *i.e*. an entity that possesses all attributes included in the given definition that, however, is "clearly" not alive or conversely, an entity that lacks some of those attributes but is "obviously" a life form, some kind of intuitive understanding of the living state superseding any definition is involved. Even corroboration of a definition of life that would involve finding more and more diverse entities fitting the definition is compromised by the same problem: to count a case as supportive, an independent criterion is required, but this can only be intuitive.

So we seem to 'know it when we see it' but defining life is an elusive goal and apparently an inherently meta-scientific (metaphysical) task. The metaphysical character of the quest for the best definition of life does not necessarily imply that the exercise is futile. On the contrary, it is easy to envisage at least two areas of utility for such definitions: *didactic* – better teaching of the fundamentals of biology and *heuristic* – formulation of new falsifiable hypotheses and perhaps identification and study of novel life forms. I will not address the didactic aspect here but will briefly discuss the potential heuristic power of life definitions after first commenting on Trifonov's approach and conclusions.

Trifonov goes about the derivation of a consensus definition of life in a manner that is unusual in scientific treatises but that has served him fairly well in previous work on the order of appearance of codons in the genetic code (4). The approach consists in compiling as many (supposedly) independent definitions as possible and then comparing vocabularies of these definitions to derive a consensus, the "essential core" that in itself may be hoped to provide for a better (the best) definition. There is no genuine scientific justification behind this approach and no guarantee that the numerous compared definitions are not all based on common misconceptions. In part, this indeed could be the case. Trifonov's core/consensus, additionally reduced through elucidation of apparent dependencies between some of the core terms, provides for a sensible, intuitively plausible "minimalist definition": *almost exact self-reproduction* or *self-reproduction with variations* (2). It is certainly interesting, as Trifonov points

Corresponding author:
Eugene V. Koonin
E-mail: koonin@ncbi.nlm.nih.gov

out, that this "objectively derived" minimalist definition of life almost exactly matches the definition given about 50 years by none other than Alexander Ivanovich Oparin, the famous even if notorious Russian biochemist who propounded the first concrete physico-chemical scenario for the origin of life.

Yet, all its simplicity and appeal notwithstanding, the minimalist definition appears to be neither necessary nor sufficient, not even internally consistent. A simple implication of information theory (and more fundamentally, thermodynamics) is that error-free replication (more precisely, any information transmission process) is impossible (5). Hence the phrase *self-reproduction with variation* is actually redundant because any replication process will be characterized by some intrinsic error rate. The problem is exactly the opposite: it has been shown by Eigen and others that for stable information transfer (inheritance) down the chain of generations to be sustained, the error rate must not exceed a certain critical value known as error catastrophe or mutational meltdown threshold (6, 7). Thus, a necessary condition for life to evolve is not simply replication and not 'replication with variation' (a tautology) but *replication with an error rate below the sustainability threshold* (Trifonov's '*almost exact self-reproduction*' fits the bill but appears imprecise). Another feature of a self-reproducing system that appears necessary for evolution is the phenotype-genotype feedback, or more precisely, differential effect of errors on the replication rate. Such differential fitness effect of mutations is a necessary condition of selection, both in its purifying and positive (Darwinian) incarnations. Hence *replication with an error rate below the sustainability threshold, with non-uniformly distributed fitness effects of errors* could be a candidate for a necessary and sufficient definition of life, or probably more precisely, a criterion for identification of life forms (5).

Both Trifonov's consensus definition and the amended one given here are strictly *informational*. Any biologist will immediately feel that "something is missing" in these definitions. In broad outline, these missing components are: i) chemistry (metabolism), ii) energy conversion, and iii) structure (various forms of compartmentalization); for brevity, these may be denoted operational components of life forms, in contrast to informational components (8).

Even more damning for the informational definitions of life, it may appear that these definitions are easily falsified by computer viruses and all forms of 'artificial life' that replicate, mutate and evolve (9) but arguably are not actual life forms. However, such falsification is illusory for the simple reason that these evolving entities themselves are produced by highly evolved life forms. Clearly, the emergence of such life forms (and human civilization in particular) was a pre-requisite for the advent of these replicators – they are not life forms as such but clearly are derived from life forms. We are currently unaware of life forms evolved from inanimate matter (or so we believe) that would lack any of the major operational attributes, in addition to replication. The operational components seem to be pre-requisite for the evolution of replicators fitting the above criteria, hence the informational definition of life presupposes the existence of some forms of metabolism, energy transformation and compartmentalization. However, all life forms on earth, with the exception of viruses which are obligate intracellular parasites, encompass a much stronger connection between replication, chemistry, energy and structure: *the replicating moiety (genome) encodes key information on the operational components (primarily in the form of proteins)*.

Is encoding operational components in the genome a necessary attribute of life? This does not appear certain at all, and here we come to the potential heuristic value of life definitions. Let us formulate a hypothesis: *there are purely informational life forms in which genomes carry only the minimal information required for replication whereas all operational components are supplied by the (conducive) environment*. Actually, hypothetical 'information-only' replicators are among the essential features of several origin of life scenarios because emergence of complex genomes encoding operational components prior to the advent of an efficient replication mechanism appears effectively impossible (5, 10). This hypothesis may be hard to falsify but it certainly can be corroborated under two types of settings: in the laboratory and in extraterrestrial biospheres if and when such are discovered. Exploration of putative life outside earth belongs in the future but vigorous attempts to create in the laboratory an evolving system of replicators employing exogenous supplies chemicals and energy are underway, and certain progress has been achieved with ribozyme polymerases (11). However, these experimental systems remain a far cry from the efficient replicators required to start the evolution of life. Thus, the jury is still out on the 'information only life' hypothesis.

A remarkable feature of all known biological replicators is their digital character: these replicators are polymers consisting of multiple types of monomers. Such polymers appear uniquely suited for information encoding, so a plausible hypothesis seems to be that *digital properties are necessary for life*. This hypothesis certainly would be falsified by the (remarkable) discovery of analog life forms.

To summarize, I believe that the 'democratic' approach applied by Trifonov to the definition of life problem did not lead him far astray and converged on a natural and sensible, if not quite precise, informational definition. In my view, although life definitions are metaphysical rather than strictly scientific propositions, they are far from being pointless and have potential to yield genuine biological insights.

References

1. E. Schroedinger. What Is Life?: with "Mind and Matter" and "Autobiographical Sketches" Cambridge University Press, Cambridge (1992).
2. E. N. Trifonov. *J Biomol Struct Dyn 29*, 259-266 (2011).

3. J. Gayon, C. Malaterre, M. Morange, F. Paulin-Cerceau, and S. Tirard. *Origins Life Evol Biospheres 40*, 119-244 (2010).
4. E. N. Trifonov. *J Biomol Struct Dyn 22*, 1-11 (2004).
5. E. V. Koonin. The Logic of Chance: The Nature and Origin of Biological Evolution FT press, Upper Saddle River, NJ (2011).
6. M. Eigen. *Naturwissenschaften 58*, 465-523 (1971).
7. J. Summers and S. Litwin. *J Virol 80*, 20-26 (2006).
8. R. Jain, M. C. Rivera, and J. A. Lake. *Proc Natl Acad Sci U S A 96*, 3801-3806 (1999).
9. C. Adami. *Nat Rev Genet 7*, 109-118 (2006).
10. E. V. Koonin. *Ann N Y Acad Sci 1178*, 47-64 (2009).
11. T. A. Lincoln and G. F. Joyce. *Science 323*, 1229-1232 (2009).