Comment

Neutral networks and chemical function in RNA
Comment on "Evolutionary dynamics of RNA-like replicator systems: A bioinformatic approach to the origin of life"

Susanna C. Manrubia

Centro de Astrobiología (INTA-CSIC), Ctra. de Ajalvir km. 4, 28850 Torrejón de Ardoz, Madrid, Spain

Understanding how complex chemical ecosystems could have emerged from simple interacting molecules is key to work out a plausible theory on the origin of life. Despite the many difficulties faced by empirical approaches to the problem [8], the conceptual framework of such a theory has been significantly developed [10] thanks to theoretical and computational studies often inspired by or relying on the properties of RNA molecules. The pioneering model of quasispecies introduced by Eigen [4] has more than guided the development of the field, conditioning our view of quasispecies dynamics through concepts such as the error threshold. In an insightful contribution [10], Takeuchi and Hogeweg begin by reviewing Eigen's model and several modifications that confer it an increasing degree of realism: The error threshold, for instance, does not depend only on the length of a sequence, but on its degree of neutrality as well. Quasispecies models have indeed tried to incorporate two important properties of RNA-like systems: the redundancy of the sequence-to-secondary structure map and, to a lesser extent, the complex topology of the sequence-to-function relationship, i.e., the fitness landscape.

The observation that an astronomically large number of RNA sequences fold into the same secondary structure [7] has major implications for short molecules, mainly, which are those of relevance in the context of the RNA world. For example, the $4^{12} \sim 1.7 \times 10^7$ possible RNA sequences of length 12 can be classified into 58 different secondary structures (or phenotypes) [1]. For the 6.9×10^{10} sequences of length 18, the number of different phenotypes raises only to 3211 [3]. The probability to access one phenotype from another is the quantity to substitute the microscopic transition probabilities between genotypes – as formulated in Eigen's original quasispecies theory. That probability depends on the skewed abundance of different phenotypes and on the topology of the corresponding neutral networks [1]. As a result, a more faithful picture of the bond between the phenotypes of a quasispecies is that of a network whose nodes represent phenotypes and whose links stand for the transition probabilities between mutually accessible nodes. Taking into account that the RNA secondary structures observed in nature typically present abundances above average [5], realistic models of quasispecies could be approximated by a core of strongly connected phenotypes. In this scenario, molecular quasispecies should be viewed as a robust, self-maintained entity with all typical phenotypes simultaneously present – unless some of them have zero fitness. The interpretation of the error threshold and the fate of the quasispecies under high enough mutation rates would have to be rethought, as well as the implications of this dynamics for

systems, as viruses, where the quasispecies theory is applied [6]. Despite the efforts to extend quasispecies theory to more realistic settings, a theory that derives mesoscopic (phenotypic) equations for RNA sequences from the microscopic (local) properties of the space of genotypes is still lacking. A subsequent improvement in modeling implies the use of realistic fitness landscapes, which can be accurately estimated only in the light of empirical data. There are correlations between molecular structure and function in RNA, for instance, which should be considered. In this context, too simplified landscapes as the peak landscape are at best poor representations of actual phenotype–function relationships. Though RNA is the most thoroughly studied model with a non-trivial genotype–phenotype map, proteins or regulatory and metabolic networks share many of its qualitative features [11]. Our steadily increasing knowledge of these systems demands that the view of genotype and phenotype as entities fulfilling a one-to-one relationship be essentially abandoned.

Models based on the phenotype are still at their infancy and, when grown up, will likely confront us with new phenomenology, as in the following example. The existence of huge neutral networks affects not only the evolution and adaptation of replicator populations, but plays a main role in the absence of replication and selection, as well. Common RNA secondary structures (those with a frequency above average) are systematically found in reasonably large populations of random RNA sequences. The sequence-to-structure-to-function redundancy offers as a side effect an alternative, plausible solution to the problem of how long molecules could have been formed in the absence of an error-correcting, template copy mechanism [2]. It has been observed that hairpin-like secondary structures represent the most abundant structural family in short RNAs [9]. At present, hairpin structures are common in viral and cellular RNAs and some of them act as ribozymes that catalyze RNA ligase reactions. These structures should have been ubiquitous in populations of random polymers and may have enhanced the ligation of short molecules. Under appropriate environmental conditions, diverse populations of long enough molecules would appear and, eventually, this process of modular evolution could have begotten new chemical functions: It is plausible that a mechanism of this kind could have ushered in the emergence of the first replicase-like molecule. At that point, a scenario dominated by RNA-like replicators, as described in [10], could have been triggered.

References

[1] Aguirre J, Buldú JM, Stich M, Manrubia SC. Topological structure of the space of phenotypes: the case of RNA neutral networks. PLoS ONE 2011;6:e26324.
[2] Briones C, Stich M, Manrubia SC. The dawn of the RNA world: toward functional complexity through ligation of random RNA oligomers. RNA 2009;15:743–9.
[3] Cowperthwaite MC, Economo EP, Harcombe WR, Miller EL, Ancel Meyers L. The ascent of the abundant: how mutational networks constrain evolution. PLoS Comp Biol 2008;4:e1000110.
[4] Eigen M. Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften 1971;58:465–523.
[5] Jörg T, Martin OC, Wagner A. Neutral network sizes of biological RNA molecules can be computed and are not atypically small. BMC Bioinform 2008;9:464.
[6] Manrubia SC, Domingo E, Lázaro E. Pathways to extinction: beyond the error threshold. Phil Trans R Soc B 2010;365:1943–52.
[7] Reidys C, Stadler PF, Schuster P. Generic properties of combinatory maps: neutral networks of RNA secondary structures. Bull Math Biol 1997;59:339–97.
[8] Schrum JP, Zhu TF, Szostak JW. The origins of cellular life. Cold Spring Harb Perspect Biol 2010;2:a002212.
[9] Stich M, Briones C, Manrubia SC. On the structural repertoire of pools of short, random RNA sequences. J Theor Biol 2008;252:750–63.
[10] Takeuchi N, Hogeweg P. Evolutionary dynamics of RNA-like replicator systems: a bioinformatic approach to the origin of life. Phys Life Rev 2012;9(3):219–63 [this issue].
[11] Wagner A. The origins of evolutionary innovations. Oxford University Press; 2012.