J Mol Evol (2013) 77:197–198
DOI 10.1007/s00239-013-9595-7

RANDOM WALKING

Evolving Wonder-RNAs in a Test Tube

Yingfu Li

Received: 8 November 2013 / Accepted: 11 November 2013 / Published online: 16 November 2013

RNA is a captivating polymer: it is chemically simple yet functionally diverse. RNA has been sought out by Mother Nature to take on many important biological functions, from genetic information encoding and transmission, to enzymatic catalysis and molecular recognition, to regulation of gene expression. Along the way, Nature has produced so many interesting RNA species, and, perhaps out of necessity, given them different sizes. At one end of the spectrum are microRNAs made of merely $\sim$22 nucleotides, and at the other end is the human 28S ribosomal RNA composed of 5,070 nucleotides! Yet, most functional RNA molecules, including many ribozymes and riboswitches, are middle-of-the-roaders having 50–200 nucleotides.

As far as the complexity of the structure and function is concerned, the size of the RNA does matter. The most fitting example is the ribosome, which is designed to execute an essential and demanding chemical transformation in a highly organized manner. Nature has adequately dealt with this challenge by creating one of most spectacular macromolecular assemblies in cells. Among several dozen macromolecules used to construct the ribosome are two super-sized RNAs that contain a few thousand nucleotides. One of Nature's most intriguing but unsolved mysteries is the evolution process behind the selection of such long functional RNA molecules. It has been postulated that the modern size of ribosomal RNAs is a consequence of amalgamation of smaller fragments (Clark 1987; Gray and Schnare 1996) and sequence comparison studies have suggested that some ribosomal RNA domains are indeed older than others (Gray and Schnare 1996; Bokov and Steinberg 2009). Evidently, Nature has found a way to patiently build a complex function through progressive, domain-by-domain construction of "wonder-RNAs".

A distinct trait of *Homo sapiens* is to learn and apply. Based on the principle of Darwinian evolution, in vitro selection technique was devised in 1990 (Ellington and Szostak 1990; Tuerk and Gold 1990), and has since been used to create numerous functional RNA (or single-stranded DNA) molecules. Key to the success of each in vitro selection experiment is the simultaneous exploration of a vast sequence space covered by a random library containing as many as 10^{15} RNA siblings. Although many fascinating functional nucleic acids have been derived by this approach, large and complex wonder-RNAs are few and far between (Bartel and Szostak 1993). This may simply be an inevitable outcome of two inherent drawbacks associated with the random library approach: the inefficient sequence space coverage and the tyranny of simple structural motifs. A standard DNA synthesis produces $\sim$0.1 mg of DNA ($\sim 10^{15}$ molecules), just enough to cover the entire sequence space of a 25-nucleotide sequence ($4^{25} \approx 10^{15}$). When the random domain increases to 50 nucleotides (still smaller than a tRNA), the sequence space rises to $\sim 10^{30}$ ($4^{50} \approx 10^{30}$), translating to $\sim$35,000 t of DNA! Thus, the sequence space coverage for a long RNA is categorically infinitesimal. Even if we assume there are complex RNAs in such a library, the chance of finding them is further diminished due to the overabundance of simple structural motifs in the library (Schlosser and Li 2005). Computational analysis has shown that random libraries are primarily populated with simple structural motifs like hairpin loops and three-way junctions and have extremely few more intricate structural motifs often found in large biological RNAs (Gevertz et al. 2005).

Y. Li (✉)
Department of Biochemistry and Biomedical Sciences, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada
e-mail: liying@mcmaster.ca

One conceivable way to access more sequence space is to use a "structured" library where a complex structural framework is engineered into the sequence of the library. Assume that a hypothetical complex functional RNA is a 50-nucleotide 4-way junction in which functionally important residues are 10 unpaired nucleotides confined by four short duplexes of five base pairs each. While more than 10^{30} (4^{50}) sequences are required to produce a specific 4-way junction, only $\sim 10^6$ (4^{10}) are required for a structured library that contains only 10 random nucleotides but four pre-engineered duplex elements (thus no need for randomization at these locations). In other words, the structured library approach will create much better sequence space coverage for the functionally important (rather than structurally important) nucleotides. An extension of this is to use a composite structured library combined from individual pools each having a different structural framework and sensibly positioned random nucleotides. To set up such a library, there are many interesting (and complex) secondary structure motifs to choose from. The design may be further benefited by taking advantage of the ever-increasing power of modern computers and proven RNA structure prediction algorithms to search for the best framework sequences that will not be significantly affected by the random nucleotides (Luo et al. 2010).

As eluded to above, it is most likely that Nature has built modern long functional RNAs step-by-step in a long evolutionary process (billions of years). It will certainly require multiple steps (and a lengthy process) to derive large and structurally complex RNAs in a test tube. Although in vitro selection technique offers a relatively quick way to isolate functional RNAs, it still takes weeks or even months to perform several rounds of selective amplification before functional species begin to dominate. If a complex RNA has to be constructed domain-by-domain, the entire process may take a really long time to complete. One way to expedite its discovery is to take advantage of the next-generation sequencing technologies that are capable of producing millions of sequence reads in a single sequencing reaction (Rothberg and Leamon 2008). Therefore, there is no need to wait for the "signal" to emerge from the evolving pool—one just needs to perform a limited number of selection cycles, send the partially enriched pool for sequencing, and pick the top-ranked sequence(s) to go to the next step (Pitt and Ferré-D'Amaré 2010). With all these wonderful tools or technologies at our disposal and surely better ones to come, it is just a matter of time until we can create large wonder-RNAs at ease, whether for pure scientific fun or for practical needs.

References

Bartel DP, Szostak JW (1993) Isolation of new ribozymes from a large pool of random sequences. Science 261:1411–1418

Bokov K, Steinberg SV (2009) A hierarchical model for evolution of 23S ribosomal RNA. Nature 457:977–980

Clark CG (1987) On the evolution of ribosomal RNA. J Mol Evol 25:343–350

Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. Nature 346:818–822

Gevertz J, Gan HH, Schlick T (2005) In vitro RNA random pools are not structurally diverse: a computational analysis. RNA 11: 853–863

Gray MW, Schnare MN (1996) Evolution of rRNA gene organization. In: Zimmerman RA, Dahlberg AE (eds) Ribosomal RNA structure, evolution, processing, and function in protein biosynthesis. CRC Press, Boca Raton, pp 49–69

Luo X, McKeague M, Pitre S, Dumontier M, Green J, Golshani A, Derosa MC, Dehne F (2010) Computational approaches toward the design of pools for the in vitro selection of complex aptamers. RNA 16:2252–2262

Pitt JN, Ferré-D'Amaré AR (2010) Rapid construction of empirical RNA fitness landscapes. Science 330:376–379

Rothberg JM, Leamon JH (2008) The development and impact of 454 sequencing. Nat Biotechnol 26:1117–1124

Schlosser K, Li Y (2005) Diverse evolutionary trajectories characterize a community of RNA-cleaving deoxyribozymes: a case study into the population dynamics of in vitro selection. J Mol Evol 61:192–206

Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249:505–510