EXPLORING CONNECTIVITY IN SEQUENCE SPACE OF FUNCTIONAL RNA. Chenyu Wei[1,2], Andrew Pohorille[1,2], Milena Popovic[2,3], and Mark Ditzler[2], [1]Department of Pharmaceutical Chemistry, UCSF; chenyu.wei@nasa.gov, [2]Exobiology Branch, NASA Ames Research Center, Mail Stop 239-4, Moffett Field, CA 94035; andrew.pohorille@nasa.gov; mark.a.ditzler@nasa.gov, [3]Blue Marble Space Institute of Science, Seattle, WA 98145; milena.popovic@nasa.gov.

Abstract: Emergence of replicable genetic molecules was one of the marking points in the origin of life, evolution of which can be conceptualized as a walk through the space of all possible sequences. A theoretical concept of fitness landscape was proposed by Wright [1-3] to help understand the evolution process, on which each genotype has a value of fitness, and the evolution of a phenotype to reach a local fitness peak is viewed through consecutive mutations. Natural selection biases evolution toward peaks of high fitness and away from valleys of low fitness [4,5], while neutral drift occurs in the sequence space without direction as mutations are introduced at random. If a landscape consists of isolated peaks there are no pathways between peaks consisting of consecutive, viable genotypes that differ by a single mutation, and evolutionary optimization is possible only through genetic recombination or alterations to the landscape due to e.g. environmental changes. But if the landscape contains large networks of neutral or near-neutral mutations, large volumes of genotypic space are crossed without marked effect on fitness, eventually chancing upon a new fitness peak [6].

Large neutral networks, especially for sufficiently long genomes, are possible or even inevitable [2,4,7, 8], the detection of which however has been elusive in experiments. While a few near-neutral evolutionary pathways have been found [9-11], recent experimental evidence indicates landscapes consist of largely isolated islands [12,13]. The generality of these results, however, is not clear, as the genome length or the fraction of functional molecules in the genotypic space might have been insufficient for the emergence of large, neutral networks. Thorough investigation on the structure of the fitness landscape is essential to understand the mechanisms of evolution available to early genomes.

RNA molecules are commonly assumed to play the pivotal role in the origin of genetic systems. They are widely believed to be early, if not the earliest, genetic and catalytic molecules, with abundant biochemical activities as aptamers and ribozymes, i.e. RNA molecules capable, respectively, to bind small molecules or catalyze chemical reactions. Here we present our recent studies on the structure of the sequence space of RNA ligase ribozymes selected through in vitro evolution. Several hundred thousands of sequences active to a different degree were obtained by way of deep sequencing and were analyzed, from which several large clusters were identified. Every sequence in a cluster can be reached from any other sequence in the same cluster through a series of other, active sequences by way of single point mutations. Sequences in a single cluster appear to adopt more than one secondary structure. The mechanism of refolding with changes of sequence in a cluster was examined. To shed light on possible evolutionary paths in the space of ribozymes, the connectivity between clusters was investigated. The length effect of a RNA molecule on the structure of its sequence space and subsequently varied evolution paths will be discussed.

References:

[1] Wright S. (1932). *Proceedings of the Sixth International Congress on Genetics.,* 355-366. [2] Gavrilets S. (2004). Fitness Landscapes and the Origin of Species (Princeton Univ. ress, Princeton). ISBN: 9780691119830. [3] Richter H. and Engelbrecht A. P. (2014). Recent Advances in the Theory and Application of Fitness Landscapes, Springer Verlag, Berlin. [4] Kauffman S.A. and Levin S. (1987). *J. Theor. Biol., 128,* 11-45. [5] Gavrilets S. (1997). *Trends Ecol. Evolut., 12,* 307-312. [6] Draghi J.A., Parsons T.L., Wagner G. P. and Plotkin J. B. (2010). *Nature 463,* 353-355. [7] Fontana W. and Schuster P. (1998). *Science 280,* 1451-1454. [8] Gravner J., Pitman D., and Gavrilets S. (2007). *J. Theor. Biol., 248,* 627-645. [9] Schultes E. A. and Bartel D. P. (2000) *Science 289,* 448-452. [10] Mandal M. and Breaker R. R. (2004). *Nat. Struct. Mol. Biol., 11,* 29-35. [11] Held, D. M., Greathouse, S. T., Agrawal, A. and Burke, D. H. (2003). *J. Mol. Evol., 57,* 299-308. [12] Jiménez, J. I., Xulvi-Brunet, R., Campbell, G. W., Turk-MacLeod, R. and Chen, I. A. (2013). *Proc. Natl. Acad. Sci. U.S.A. 110,* 14984-14989. [13] Petrie, K. L. and Joyce, G. F. (2014) J. *Mol. Evol., 79,* 75-90.