Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

Experimental fitness landscapes to understand the molecular evolution of RNA-based life

Shreyas S Athavale[1], Brad Spicer[1,2] and Irene A Chen[1,2]

In evolutionary biology, the relationship between genotype and Darwinian fitness is known as a fitness landscape. These landscapes underlie natural selection, so understanding them would greatly improve quantitative prediction of evolutionary outcomes, guiding the development of synthetic living systems. However, the structure of fitness landscapes is essentially unknown. Our ability to experimentally probe these landscapes is physically limited by the number of different sequences that can be identified. This number has increased dramatically in the last several years, leading to qualitatively new investigations. Several approaches to illuminate fitness landscapes are possible, ranging from tight focus on a single peak to random speckling or even comprehensive coverage of an entire landscape. We discuss recent experimental studies of fitness landscapes, with a special focus on functional RNA, an important system for both synthetic cells and the origin of life.

Addresses

[1] Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA 93106, United States

[2] Program in Biomolecular Sciences and Engineering, University of California, Santa Barbara, CA 93106, United States

Corresponding author: Chen, Irene A (chen@chem.ucsb.edu)

Current Opinion in Chemical Biology 2014, **22**:35–39

This review comes from a themed issue on **Synthetic biology**

Edited by **Pier Luigi Luisi**, **Pasquale Stano** and **Cristiano Chiarabelli**

http://dx.doi.org/10.1016/j.cbpa.2014.09.008

1367-5931/© 2014 Published by Elsevier Ltd.

Introduction

In evolutionary biology, fitness refers to the ability of organisms to survive and reproduce in their environment. In the space of all possible sequences, the mapping of genotypes to their fitness values is called a fitness landscape (also called an adaptive or evolutionary landscape). A peak in a fitness landscape therefore represents a family of highly fit genotypes that share sequence similarity. The idea of a fitness landscape was first proposed by Wright [1], who conceived of a landscape of population fitness in the space of allele frequencies. In 1970, Smith put forth the idea of evolution in a protein space, in which neighboring sequences differ by a single amino-acid substitution, leading to the modern notion of a fitness landscape in sequence space [2]. On such a landscape, the process of evolution is formally equivalent to a random walk with a bias toward hill-climbing [3,4]. Therefore, a true understanding of evolution requires knowledge of the structure of fitness landscapes. Several models of landscape topology have been analyzed [5], but the relevance of such models to real systems is unclear. Since most experiments can sample only an infinitesimal fraction of the astronomical number of possible genotypes, little is known about the overall structures of fitness landscapes. Efforts to infer landscape properties based on limited samples of sequence space can demonstrate short-term evolutionary trajectories, as discussed below, but they may lead to incorrect inferences about evolutionary dynamics [6]. The literature on fitness landscapes in general is large and has been reviewed elsewhere [7,8]. In this brief review, we focus on recent empirical investigation of fitness landscapes for functional RNA, which is now advancing rapidly due to the availability of high-throughput sequencing.

Comprehensive investigation of fitness landscapes for modern organisms is experimentally intractable, as the amount of material needed to synthesize each possible sequence is vast (e.g., for the smallest known genome, an RNA viroid, $4^{246} \approx 10^{148}$ possible sequences $\approx 10^{66}$ times the number of elementary particles in the universe) [9]. Therefore, experiments along this vein have focused primarily on a smaller, simpler entity: functional RNA. Life is believed to have originated using RNA-based genetic and metabolic systems [10–13]. RNA can store and pass on genetic information, as seen in RNA viruses. Beginning in the 1990s, RNAs have been discovered through *in vitro* evolution to catalyze a variety of fundamental biochemical reactions, such as DNA/RNA cleavage, RNA splicing and ligation, DNA/RNA phosphorylation, RNA aminoacylation, amide bond formation and cleavage and glycosidic bond formation [14–18]. The evolution and engineering of RNAs having particular binding or catalytic properties has developed into a high art, resulting in RNAs that transcribe ribozymes [19], self-replicate [20], and form cooperative networks [21], supporting the idea that an RNA-based synthetic cell may not be far off [22]. While most ribozymes are typically in the range of 30–200 bases long, some are quite short [23], such that a comprehensive, experimental investigation of their sequence space is indeed possible in principle. Shorter RNAs are of particular interest for the origin of life because the yield of polymerization and/or ligation chemistries would decrease as length increases

[24], and because specific short sequence motifs are correspondingly more common than long motifs (the 'tyranny of short motifs' [25]). Functional RNA is thus a tractable model system for comprehensive studies of evolution, particularly as it relates to early or synthetic life.

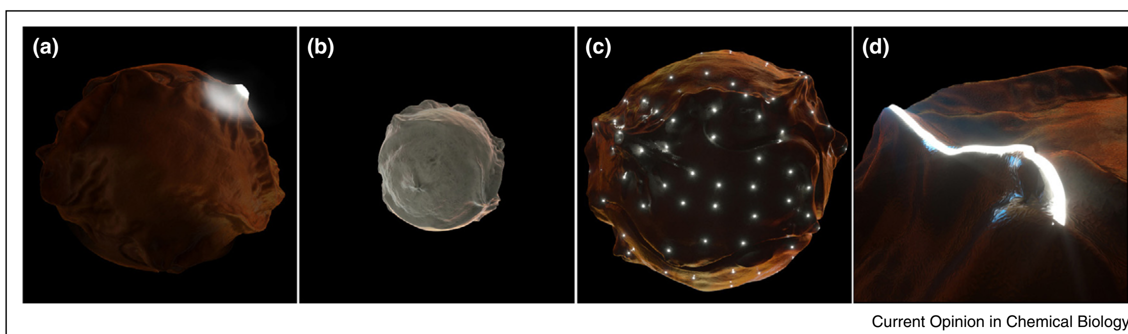
The impact of high-throughput sequencing

With the advent of high-throughput sequencing (HTS), it became possible to rapidly generate large amounts of sequence data at very low per-base cost. On the widely used Illumina platform, obtaining 10^7 sequence reads is now routine, with 10^9 reads possible at the high end. Taking advantage of these advances, Pitt and Ferre d'Amare first demonstrated the utility of HTS for analysis of functional RNA. They mutagenized the 54 nt class II RNA ligase ribozyme, originally discovered by *in vitro* evolution, and reselected the variant pool for active ligases, thereby illuminating the fitness landscape around the original ligase sequence (Figure 1a) [26]. Using several orders of magnitude more sequence information than what had been possible with Sanger sequencing, this study quickly identified residues important for maximal ribozyme activity. HTS has also been used to discover RNA aptamers for therapeutic and diagnostic purposes. For example, the *in vitro* evolution of RNAs binding to HIV-1 reverse transcriptase (RT) was followed by HTS [27]. The depth of resolution permitted identification of an asymmetric internal loop structure that inhibited RT activity and HIV replication more potently than any previously identified motifs. Increased depth of resolution from HTS also permits earlier detection of aptamers, potentially reducing the number of labor-intensive selection rounds [28,29]. The power of HTS for quantifying rare events is demonstrated in its use to characterize the results of mutagenic treatment [30]. Indeed, the *in vitro* selection of DNA aptamers containing an additional unnatural 'xeno' nucleobase pair (Z:P) was made possible by a sequencing strategy that relied upon quantitative analysis of the informational conversion of Z and P to standard bases [31••]. HTS is rapidly transforming the field of *in vitro* evolution, enabling qualitatively new inquiries.

Complete maps of fitness landscapes

The true structure of fitness landscapes can now be addressed using HTS. A major hurdle for experimental landscapes is the exponential growth of diversity as sequence length increases, which places hard limits on the maximum length that can be comprehensively interrogated in the laboratory. In addition, the sequencing output, even with HTS, limits the depth to which the landscape can be characterized. Therefore, most experimental landscapes that have been studied cover a fraction of sequence space, either at random or in the vicinity of a known wild type, making rigorous evolutionary inferences difficult. However, the potential for fundamental evolutionary insights drives a quest for complete maps of fitness landscapes. For example, computational work on RNA secondary structure suggested that sequences that fold into the same structure are common, and that this landscape of shapes resembles a network of neutral evolutionary pathways that join sequences with identical secondary structures [32,33]. From this, one might infer that, if the architecture of fitness landscapes for functional activity mirrors that for shapes, populations would readily drift through sequence space along these neutral networks. As a consequence, mutations that increase fitness should be easily accessible during evolution, ultimately resulting in discovery of the global optimum. On the other hand, some studies on experimental

Figure 1



Illumination of fitness landscapes in sequence space through different experimental approaches. **(a)** *Spotlight on a peak*: Mutagenesis of a known functional sequence and reselection of active molecules from the variant pool results in the illumination of the fitness landscape around the original active sequence (for example [26]). **(b)** *Total illumination*: Complete randomization at all positions of a relatively short sequence and selection for active molecules results in illumination of the entire landscape (for example [36••]). For this approach to work, the number of dimensions in sequence space must be relatively small. **(c)** *Random speckling*: Selection of functional biopolymers from random sequences results in many pinpoints of illumination randomly distributed over the entire fitness landscape (for example [46]). The resulting illumination is unbiased by the historical course of evolution. **(d)** *Hiker's path*: Directed or experimental evolution with specific start and end locations results in the discovery of one or more high-fitness pathways in the landscape (for example [49]).

evolution of organisms, while not comprehensive, suggest that real fitness landscapes may be quite rugged (e.g., due to epistasis), such that initial conditions would play a large role in determining evolutionary outcomes (for example, see Refs. [34,35]). Complete, experimental maps of fitness landscapes would give quantitative answers to such issues (Figure 1B).

We recently reported such a map in the space of 24-mer RNAs [36••]. Nearly all possible 24-mers were subjected to a selection for interaction with GTP covalently linked to an agarose resin, and HTS identified sequences meeting the selection criterion. Roughly 15 fitness peaks were identified. In general, peaks were essentially evolutionarily isolated from one another, in that fitness could not be preserved along any path between two peaks. A few viable evolutionary pathways were identified, but such pathways were uncommon and were more comparable to constrained paths rather than to a neutral network. As observed from the single-peak study of the class II ligase [26] and some proteins (e.g., Ref. [37]), fitness tended to drop off quickly from a peak summit, such that a random mutational excursion to another peak is expected to be rare. One might therefore speculate that initial conditions would play a determining role in the evolution of early life, but several caveats must be kept in mind. For example, only a single selection process was studied, so broader conclusions are difficult to draw. Gene duplication might enable exploration of sequence space without fitness cost. Nevertheless, this type of investigation opens the door to insights about chance, reproducibility, and optimality during the origin of life.

Might neutral evolutionary networks exist under other circumstances? Lowering the selection stringency (lowering the 'sea level') could allow less fit sequences to join evolutionary pathways. More broadly, a major question is how landscapes as a whole change as environments change. Studies at the organismal level show that environmental changes alter fitness landscapes [38]. Indeed, different environments can uncover new activities in RNA. The RNA World presumably flourished in an anoxic environment with abundant soluble iron (Fe^{2+}). Recent evidence shows that Fe^{2+} can substitute for Mg^{2+} in RNA structure and function [39]. In addition, Fe^{2+} expands the catalytic repertoire of RNA, as structured RNAs in complex with Fe^{2+} can catalyze electron transfer reactions [40••]. Perhaps environmental changes are a key to accessing neutral networks of mutations or new fitness peaks corresponding to evolutionary innovations.

Outside the RNA World

Given the presumed importance of RNA for the origin of life, the first glimpse into a comprehensive fitness landscape for RNA begs the question: how do the landscape architectures of RNA compare with other biopolymers? If life were to start with an alternative biopolymer, would its evolution differ in important ways? A robust line of inquiry into the functional potential of a number of non-biological nucleic acids has emerged over the past several years. These evolvable systems include modifications of the backbone as well as additional bases, and have been the subject of recent review in this journal [41]. More exotic alternative nucleic acids, inspired by the search for prebiotic precursors to RNA, also exist [42•,43]. In addition, *in vitro* evolution is likely to expand in the near future beyond nucleic acids and proteins [44••]. Systematic comparison of the functional potential, and evolutionary landscape architectures, of these alternative systems is likely to be on the horizon.

At this time, the only other biopolymer that has been investigated in systematic detail is proteins. Again, only a small number of variable sites can be fully investigated (e.g., nearly a kilogram of material is required to contain at least one molecule of each possible 18-mer peptide), so this saturation analysis has been limited to variations within a known protein. In particular, β-lactamase has been the target of extensive mutagenesis studies. Recently, Jacquier *et al.* characterized most possible single DNA mutations of the β-lactamase gene TEM-1, representing roughly a fifth of possible single amino acid changes [45•]. The distribution of fitness effects in these mutants was bimodal, with most mutations having little effect on the minimum inhibitory concentration for amoxicillin, and a minority (13%) of mutations essentially inactivating the enzyme.

While such studies shine a bright light on the immediate neighborhood of the wild-type sequence, the opposite approach is also informative. *In vitro* or *in vivo* selection of functional proteins from random sequence by several techniques allows pinpoint illumination over the entire fitness landscape, unbiased by the history of life so far (Figure 1C) [46]. On the other hand, 'shuffling' approaches take advantage of local covariation of sites within known functional sequences to bias the search toward sequences more likely to retain function [47], an approach widely used in protein evolution and engineering. When recently applied to ribozymes, synthetic shuffling and *in vitro* selection based on a known kinase ribozyme resulted in a substantially improved recombinant ribozyme [48]. Although these 'speckled' illumination approaches have only explored a vanishingly small fraction of possible sequence space thus far, one could envision defining the entire landscape for an activity appropriate for short peptides.

Another approach focuses on the discovery of specific evolutionary pathways from a given starting point to a new fitness peak (Figure 1D). This approach is usually termed directed evolution when a specific gene is to be

optimized for a particular application, or experimental evolution when the entire organismal genome is permitted to evolve. In general, many studies show that such pathways can indeed be found. For example, an arbitrarily chosen sequence evolved the ability to rescue defective phage by enhancing infectivity by several orders of magnitude [49]. *Escherichia coli* propagated in the laboratory for more than 30,000 generations evolved the ability to utilize citrate as a carbon source under oxic conditions [50]. A single point mutation alters the specificity of an aptamer from guanine to adenine [51]. A novel protein fold evolved from an old one during functional optimization [52•]. Such studies illustrate the striking breadth of evolutionary potential, although it is difficult to draw quantitative conclusions about the fitness landscape from them. Nevertheless, these studies vividly demonstrate biological creativity during evolution and serve the practical purpose of developing new functional sequences for synthetic organisms.

Conclusion

In this brief review, we have taken the perspective that studies of fitness landscapes build our fundamental understanding of evolution. However, the same data, taken for application-oriented functions, can be interpreted as sequence-activity relationships. Covariations have already been exploited to inform and deduce structural models of RNA [26,53]. Perhaps greater biophysical insights lie buried within deep sequencing data, awaiting discovery.

Acknowledgements

We are grateful for support from the Simons Foundation (grant no. 290356), the Foundational Questions in Evolutionary Biology Fund (grant no. RFP-12-05), and the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred. The funding entities had no role in study design, collection, analysis or interpretation of data, the writing of the report, or the decision to submit the article for publication. We also thank Peter Allen for making the illustration.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Wright S: **The roles of mutation, inbreeding, crossbreeding and selection in evolution**. *Proc VIth Int Congress of Genetics, vol. 1*. 1932:355-366.
2. Smith JM: **Natural selection and the concept of a protein space**. *Nature* 1970, **225**:563-564.
3. Kauffman S, Levin S: **Towards a general theory of adaptive walks on rugged landscapes**. *J Theor Biol* 1987, **128**:11-45.
4. van Nimwegen E, Crutchfield JP, Huynen M: **Neutral evolution of mutational robustness**. *Proc Natl Acad Sci U S A* 1999, **96**:9716-9720.
5. Orr HA: **The genetic theory of adaptation: a brief history**. *Nat Rev Genet* 2005, **6**:119-127.
6. Otwinowski J, Plotkin JB: **Inferring fitness landscapes by regression produces biased estimates of epistasis**. *Proc Natl Acad Sci U S A* 2014. (Epub ahead of print).
7. de Visser JA, Krug J: **Empirical fitness landscapes and the predictability of evolution**. *Nat Rev Genet* 2014. (Epub ahead of print).
8. Kogenaru M, de Vos MG, Tans SJ: **Revealing evolutionary pathways by fitness landscape reconstruction**. *Crit Rev Biochem Mol Biol* 2009, **44**:169-174.
9. Rocheleau L, Pelchat M: **The subviral RNA database: a toolbox for viroids, the hepatitis delta virus and satellite RNAs research**. *BMC Microbiol* 2006, **6**:24.
10. Crick FH: **The origin of the genetic code**. *J Mol Biol* 1968, **38**:367-379.
11. Orgel LE: **Evolution of the genetic apparatus**. *J Mol Biol* 1968, **38**:381-393.
12. Woese C: *The Genetic Code*. New York: Harper and Row; 1967, .
13. Gilbert W: **The RNA world**. *Nature* 1986, **319**:618.
14. Weigand BS, Zerressen A, Schlatterer JC, Helm M, Jäschke A: **Catalytically active RNA molecules: tools in organic chemistry**. In *The Aptamer Handbook.* Edited by Klussman S. Wiley-VCH; 2006:211-227.
15. Moretti JE, Müller UF: **A ribozyme that triphosphorylates RNA 5′-hydroxyl groups**. *Nucleic Acids Res* 2014, **42**:4767-4778.
16. Tuerk C, Gold L: **Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase**. *Science* 1990, **249**:505-510.
17. Ellington AD, Szostak JW: **In vitro selection of RNA molecules that bind specific ligands**. *Nature* 1990:346.
18. Robertson DL, Joyce GF: **Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA**. *Nature* 1990, **344**:467-468.
19. Wochner A, Attwater J, Coulson A, Holliger P: **Ribozyme-catalyzed transcription of an active ribozyme**. *Science* 2011, **332**:209-212.
20. Robertson MP, Joyce GF: **Highly efficient self-replicating RNA enzymes**. *Chem Biol* 2014, **21**:238-245.
21. Vaidya N, Manapat ML, Chen IA, Xulvi-Brunet R, Hayden EJ, Lehman N: **Spontaneous network formation among cooperative RNA replicators**. *Nature* 2012, **491**:72-77.
22. Blain JC, Szostak JW: **Progress toward synthetic cells**. *Annu Rev Biochem* 2014, **83**:615-640.
23. Yarus M: **The meaning of a minuscule ribozyme**. *Philos Trans R Soc Lond B Biol Sci* 2011, **366**:2902-2909.
24. Derr J, Manapat ML, Rajamani S, Leu K, Xulvi-Brunet R, Joseph I, Nowak MA, Chen IA: **Prebiotically plausible mechanisms increase compositional diversity of nucleic acid sequences**. *Nucleic Acids Res* 2012, **40**:4711-4722.
25. Ellington AD, Chen X, Robertson M, Syrett A: **Evolutionary origins and directed evolution of RNA**. *Int J Biochem Cell Biol* 2009, **41**:254-265.
26. Pitt JN, Ferré-D'Amaré AR: **Rapid construction of empirical RNA fitness landscapes**. *Science* 2010, **330**:376-379.
27. Ditzler MA, Lange MJ, Bose D, Bottoms CA, Virkler KF, Sawyer AW, Whatley AS, Spollen W, Givan SA, Burke DH: **High-throughput sequence analysis reveals structural diversity and improved potency among RNA inhibitors of HIV reverse transcriptase**. *Nucleic Acids Res* 2013, **41**:1873-1884.
28. Bawazer LA, Newman AM, Gu Q, Ibish A, Arcila M, Cooper JB, Meldrum FC, Morse DE: **Efficient selection of biomineralizing DNA aptamers using deep sequencing and population clustering**. *ACS Nano* 2014, **8**:387-395.
29. Cho M, Soo OS, Nie J, Stewart R, Eisenstein M, Chambers J, Marth JD, Walker F, Thomson JA, Soh HT: **Quantitative selection

and parallel characterization of aptamers. *Proc Natl Acad Sci U S A* 2013, **110**:18460-18465.

30. Petrie KL, Joyce GF: **Deep sequencing analysis of mutations resulting from the incorporation of dNTP analogs**. *Nucleic Acids Res* 2010, **38**:8095-8104.

31. Sefah K, Yang Z, Bradley KM, Hoshika S, Jiménez E, Zhang L,
•• Zhu G, Shanker S, Yu F, Turek D *et al.*: **In vitro selection with artificial expanded genetic information systems**. *Proc Natl Acad Sci U S A* 2014, **111**:1449-1454.

The authors apply *in vitro* evolution to an artificially expanded genetic information system (AEGIS, which includes the nonstandard base pair Z:P). They identify DNA aptamers, built from 6 different nucleotides, capable of binding to breast cancer cells. The enriched GACTZP DNA apatamers were deep sequenced by copying Z and P nucleotides into standard DNA using 'conversion' protocols.

32. Schuster P, Stadler PF: **Landscapes: complex optimization problems and biopolymer structures**. *Comput Chem* 1994, **18**:295-325.

33. Schuster P, Fontana W, Stadler PF, Hofacker IL: **From sequences to shapes and back: a case study in RNA secondary structures**. *Proc Biol Sci* 1994, **255**:279-284.

34. Kouyos RD, Leventhal GE, Hinkley T, Haddad M, Whitcomb JM, Petropoulos CJ, Bonhoeffer S: **Exploring the complexity of the HIV-1 fitness landscape**. *PLoS Genet* 2012, **8**:e1002551.

35. Kvitek DJ, Sherlock G: **Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape**. *PLoS Genet* 2011, **7**:e1002056.

36. Jimenez JI, Xulvi-Brunet R, Campbell G, Turk-MacLeod R,
•• Chen IA: **Comprehensive experimental fitness landscape and evolutionary network for small RNA**. *Proc Natl Acad Sci U S A* 2013, **110**:14984-14989.

This is the first study to report a complete fitness landscape, which is possible for short RNAs. The landscape was largely composed of peaks that were isolated from each other, suggesting an important role for chance events.

37. Firnberg E, Labonte JW, Gray JJ, Ostermeier M: **A comprehensive, high-resolution map of a gene's fitness landscape**. *Mol Biol Evol* 2014, **31**:1593-1605.

38. Flynn KM, Cooper TF, Moore F-G, Cooper VS: **The environment affects epistatic interactions to alter the topology of an empirical fitness landscape**. *PLoS Genet* 2013, **9**:e1003426.

39. Athavale SS, Petrov AS, Hsiao H, Watkins D, Prickett CD, Gossett JJ, Lie L, Bowman JC, O'Neill E, Bernier CR *et al.*: **RNA folding and catalysis mediated by iron (II)**. *PLoS ONE* 2012, **7**:e38024.

40. Hsiao C, Chou I, Okafor CD, Bowman JC, O'Neill EB, Athavale SS,
•• Petrov AS, Hud NV, Wartell RM, Harvey SC *et al.*: **RNA with iron (II) as a cofactor catalyses electron transfer**. *Nat Chem* 2013, **5**:525-528.

The authors show that a simple environmental change, the use of Fe^{2+} instead of Mg^{2+} in ribozymes, uncovers new catalytic activity. RNAs with well-defined 3D coordination geometries with the divalent ions catalyzed electron transfer reactions.

41. Pinheiro VB, Holliger P: **The XNA world: progress towards replication and evolution of synthetic genetic polymers**. *Curr Opin Chem Biol* 2012, **16**:245-252.

42. Chen MC, Cafferty BJ, Mamajanov I, Gállego I, Khanam J,
• Krishnamurthy R, Hud NV: **Spontaneous prebiotic formation of a beta-ribofuranoside that self-assembles with a complementary heterocycle**. *J Am Chem Soc* 2014, **136**:5640-5646.

The authors report the prebiotic synthesis of β-furanosyl nucleosides in high yields upon multiple cycles of mild heating and rehydration. These nucleosides can self-assemble into micrometer-length fibers composed of thousands of ordered structures, suggestive of a primitive system of noncovalent information storage.

43. Krishnamurthy R: **RNA as an emergent entity: an understanding gained through studying its non-functional alternatives**. *Synlett* 2014, **25** A–G.

44. Niu J, Hili R, Liu DR: **Enzyme-free translation of DNA into**
•• **sequence-defined synthetic polymers structurally unrelated to nucleic acids**. *Nat Chem* 2013, **5**:282-292.

The authors report a method to translate DNA templates into specific sequences of unnatural biopolymers. This 'reboot' of translation could lead to *in vitro* evolution of functional molecules made from a greatly expanded set of organic precursors.

45. Jacquier H, Birgy A, Le Nagard H, Mechulam Y, Schmitt E, Glodt J,
• Bercot B, Petit E, Poulain J, Barnaud G *et al.*: **Capturing the mutational landscape of the beta-lactamase TEM-1**. *Proc Natl Acad Sci U S A* 2013, **110**:13067-13072.

The authors generated a 'spotlight' mutational landscape for beta-lactamase TEM-1 using random mutagenesis, advancing a significant line of inquiry on the fitness landscape of this enzyme.

46. Golynskiy MV 3rd, Haugner JC, Morelli A, Morrone D, Seelig B: **In vitro evolution of enzymes**. *Methods Mol Biol* 2013, **978**:73-92.

47. Stemmer WPC: **Rapid evolution of a protein in vitro by DNA shuffling**. *Nature* 1994, **370**:389-391.

48. Curtis EA, Bartel DP: **Synthetic shuffling and in vitro selection reveal the rugged adaptive fitness landscape of a kinase ribozyme**. *RNA* 2013, **19**:1-13.

49. Hayashi Y, Sakata H, Makino Y, Urabe I, Yomo T: **Can an arbitrary sequence evolve towards acquiring a biological function?** *J Mol Evol* 2003, **56**:162-168.

50. Blount ZD, Borland CZ, Lenski RE: **Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli***. *Proc Natl Acad Sci U S A* 2008, **105**:7899-7906.

51. Mandal M, Breaker RR: **Adenine riboswitches and gene activation by disruption of a transcription terminator**. *Nat Struct Mol Biol* 2004, **11**:29-35.

52. Chao FA, Morelli A, Haugner JC III, Churchfield L, Hagmann LN,
• Shi L, Masterson LR, Sarangi R, Veglia G, Seelig B: **Structure and dynamics of a primordial catalytic fold generated by in vitro evolution**. *Nat Chem Biol* 2013, **9**:81-83.

In vitro evolution of a small non-catalytic zinc finger protein resulted in the discovery of a new RNA ligase enzyme. Remarkably, the new enzyme lost the original zinc finger scaffold and exhibits a highly flexible, novel protein fold.

53. Lucks JB, Mortimer SA, Trapnell C, Luo S, Aviran S, Schroth GP, Pachter L, Doudna JA, Arkin AP: **Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq)**. *Proc Natl Acad Sci U S A* 2011, **108**:11063-11068.