Annual Review of Biophysics

Molecular Fitness Landscapes from High-Coverage Sequence Profiling

Celia Blanco,[1] Evan Janzen,[1,2,*] Abe Pressman,[1,3,*] Ranajay Saha,[1] and Irene A. Chen[2]

[1]Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106, USA; email: blanco@ucsb.edu, evanjanzen@ucsb.edu, rsaha@ucsb.edu, ichen@ucsb.edu, abe_pressman@ucsb.edu

[2]Biomolecular Science and Engineering Program, University of California, Santa Barbara, California 93106, USA

[3]Department of Chemical Engineering, University of California, Santa Barbara, California 93106, USA

Annu. Rev. Biophys. 2019. 48:1–18

First published as a Review in Advance on January 2, 2019

The *Annual Review of Biophysics* is online at biophys.annualreviews.org

https://doi.org/10.1146/annurev-biophys-052118-115333

*These authors contributed equally to this article

Keywords

fitness landscape, high-throughput sequencing, ribozyme, deep mutational scanning

Abstract

The function of fitness (or molecular activity) in the space of all possible sequences is known as the fitness landscape. Evolution is a random walk on the fitness landscape, with a bias toward climbing hills. Mapping the topography of real fitness landscapes is fundamental to understanding evolution, but previous efforts were hampered by the difficulty of obtaining large, quantitative data sets. The accessibility of high-throughput sequencing (HTS) has transformed this study, enabling large-scale enumeration of fitness for many mutants and even complete sequence spaces in some cases. We review the progress of high-throughput studies in mapping molecular fitness landscapes, both in vitro and in vivo, as well as opportunities for future research. Such studies are rapidly growing in number. HTS is expected to have a profound effect on the understanding of real molecular fitness landscapes.

Contents

1. INTRODUCTION

Predicting evolution is a key challenge in biological science that not only tests our basic understanding but also has real-world ramifications. For example, prediction of influenza virus evolution (61) is used to select vaccine strains. In principle, evolutionary trajectories could be predicted probabilistically if one knew how any mutation would affect the fitness of the organism or molecule (as well as knowing other parameters, including population size and mutation rate). The function of fitness in sequence space is known as the fitness landscape (98, 113). Evolution can be seen as a random walk (i.e., exploration by mutation) on a fitness landscape with a bias toward hill climbing (i.e., selection for higher fitness) (50). Despite the importance of mapping fitness landscapes, the size of sequence space is astronomically large (m^N points for an alphabet size m and sequence of length N), which has previously hampered substantial mapping efforts. While experiments in the laboratory can include a large number of biopolymer sequences (e.g., up to 10^{17} molecules for in vitro evolution of RNA), analysis is also limited by sequencing capacity. Therefore, within the last decade, analysis has been transformed by the accessibility of high-throughput sequencing (HTS), as fitness data can now be collected on millions of sequences in parallel. These data form a quantitative framework for addressing classic questions: How does the topography of the fitness landscape constrain evolution? How repeatable are evolutionary outcomes? What does the topography teach us about the emergence of new structures and functions?

In this review, we highlight progress that has been made to map fitness landscapes empirically using high-throughput techniques, focusing on biomolecules. To give an initial context for these studies, we first introduce simple models of fitness landscapes and their properties. Next, we consider the case study of a classic question, how well selection can optimize fitness on real landscapes, and the impact of HTS on this problem. We then devote our attention to other ways in which HTS has deepened our understanding of molecular fitness landscapes, where fitness approximates functional activity. Finally, we consider organismal fitness landscapes and the importance of the environment, a combination that is daunting in scope but the source of Darwin's "endless forms most beautiful."

Fitness: quantitative measure of evolutionary favorability, corresponding to reproduction and survival in vivo; can be defined in multiple ways in vitro

Sequence space: the set of m^N possible sequences with N variable positions, given an alphabet size of m

Fitness landscape: function of fitness over sequence space

Alphabet: the set of chemical monomers ("letters") used in the construction of a biopolymer such as DNA or protein

High-throughput sequencing (HTS): sequencing technology that reads 10^7–10^{12} bases of DNA; platforms include Illumina, 454, PacBio, Oxford Nanopore, ABI-SOLiD, and others

2. SEQUENCE SPACE

Sequence space is discrete, where the number of dimensions N is equal to the number of variable monomer sites in a biopolymer (e.g., with no fixed sites, N is the sequence length), and the number of points in each dimension is the alphabet size m. Fitness is a continuous variable that describes a sequence's evolutionary favorability and can be defined depending on experimental context. Plotting fitness over sequence space gives the fitness landscape of $N + 1$ dimensions. To gain insight, one may consider drawings of the space of very small binary sequences, with fitness represented as a heat map (**Figure 1**).

For standard RNA or DNA, with an alphabet size of four nucleotides, the size of sequence space is 4^N ($\approx 10^{0.6N}$). The amount of nucleic acid one might work with in vitro would typically be $<10^{17}$ molecules, so sequence space becomes experimentally intractable in the lab for $N > 27$ if one desires full coverage of the space. For standard proteins, composed of 20 amino acids, the space 20^N ($\approx 10^{1.3N}$) becomes intractable in vitro for $N > 12$ at full coverage. For experimental evolution in vivo (e.g., in microbes), a 1-L experiment might contain 10^{12} cells, allowing up to $\sim$20 genome sites to be covered in full. In practice, fitness landscapes can be fully mapped for relatively short sequences, while fitness landscapes for organisms and larger molecules must focus on a small number of variable sites or sparsely sample the sequence space.

Although sequence space is exponentially large, it is still a special subset of the larger space of all possible chemicals. Sequence space for a particular polymer type (biological or artificial) can be thought of as a sort of filigree in chemical space, defined by its particular bonding patterns, which is closely apposed to those for similar polymer types (74).

3. SIMPLE MODELS OF FITNESS LANDSCAPES

Experimental investigation of fitness landscapes is difficult owing to the complexity of sequence space, so a substantial body of work has involved the development of theoretical models of fitness landscapes. These models can be applied to biological data as a way to represent complex patterns with a small number of parameters. Although theoretical models for fitness landscapes have been reviewed elsewhere (17, 66), we introduce here two simple and influential models [Mount Fuji and *NK* (number of sites, N, and the epistatic degree, K)] and related models (rough Mount Fuji

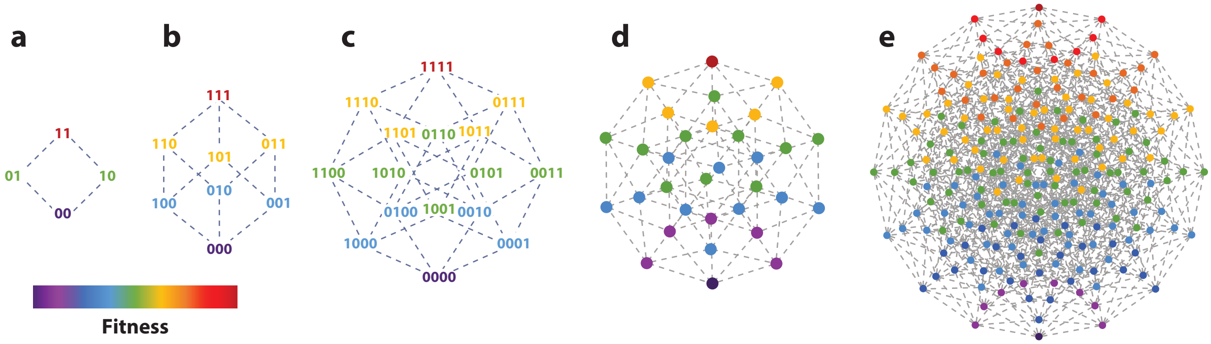

Figure 1

Mock fitness landscapes of small binary sequences, depicted as a projection of the N-dimensional hypercube. Landscapes are drawn with $m = 2$ and (*a*) $N = 2$, (*b*) $N = 3$, (*c*) $N = 4$, (*d*) $N = 5$, and (*e*) $N = 8$. The fitness of each point in sequence space is represented by color (see legend) according to a smooth so-called Mount Fuji landscape (e.g., fitness related to the number of 1s). As N increases, the number of points and neighbors increases exponentially, making a full representation of the fitness landscape difficult to interpret at higher N. Figure adapted from Reference 113. Abbreviations: m, alphabet size; N, number of dimensions (i.e., sequence length).

Epistasis: interaction of sites where a mutation's fitness contribution depends on genetic background (e.g., difference between observed fitness and additive expectation)

Reciprocal sign epistasis: sign epistasis in which mutations that are separately advantageous became jointly unfavorable (or vice versa)

Ruggedness: property describing an epistatic, uncorrelated fitness landscape of many local peaks and valleys; can be estimated quantitatively in multiple ways

Neutral mutation: mutation with little or no effect on fitness

and house of cards) to develop some understanding of possible topographies and their possible mechanisms of origin.

The simplest theoretical model is the Mount Fuji landscape (3), named after Japan's highest mountain because it is a smooth, single-peak landscape. Mount Fuji landscapes are defined as those in which every point on the sequence space—other than the global optimum—has at least one neighbor sequence (one mutational step away) of higher fitness. The simplest Mount Fuji model corresponds to a perfectly smooth, monotonic climb along any path toward the center. This topography can be created if the effect of individual mutations is additive [the effect of each site does not depend on the others (i.e., there is no epistasis)]. The absence of local optima on Mount Fuji–type landscapes allows good reconstruction of the topography even when incomplete random sampling is performed. Under conditions of strong selection and weak mutation (SSWM) (38), evolution on Mount Fuji–type landscapes results in the optimal sequence.

Most empirical landscapes exhibit certain epistatic interactions that the Mount Fuji model cannot emulate. In particular, Mount Fuji–type landscapes cannot describe reciprocal sign epistasis, in which the presence of one mutation *a* changes if another mutation *b* is beneficial, and vice versa, creating multiple optima (78). These nonadditive effects disrupt the smoothness of a landscape, creating a need for models with tunable ruggedness. A popular model of this type is the *NK* landscape (50, 51), in which the system can be solely described by two parameters: the number of sites, N, and the epistatic degree (the number of other sites influencing the effect of a given site), K. When $K = 0$, the *NK* model gives a Mount Fuji landscape. As K increases, the ruggedness of the landscape increases and local optima arise, although a global optimum is still present. In its most rugged incarnation ($K = N - 1$), the fitness contribution of a single position is affected by mutations at every other position in the sequence. In this case, the landscape is dominated by high-order epistasis, leading to a completely uncorrelated landscape with an average number of local optima $[2^N / (N - 1)]$ that scales roughly exponentially with N (**Figure 2**). A landscape in which the fitnesses of related sequences are totally uncorrelated is also known as the random house of cards model, because pulling a card (i.e., a mutation) from the house results in its collapse (i.e., complete change of the fitness landscape); the house then needs to be entirely rebuilt from a reshuffled genomic deck (53). Although interesting as a theoretical limit, the completely uncorrelated landscape probably does not occur in reality. Whether incomplete sampling of sequence space can result in a reasonable representation of the topography depends on the ruggedness of the landscape and the properties to be analyzed.

Two modifications to the *NK* model can be introduced to increase its realism. First, since proteins are often modular (e.g., composed of independent domains), the *NK* model can be adapted to include different degrees of correlation on the landscape (70). In the block (or domains) model, mutations in one block affect only the contribution of that block to the overall fitness of the protein, and each independent block can have different values of K. Blocks need not correspond to structural domains from the primary sequence but could represent amino acids that interact in the protein's tertiary structure. Second, although the original *NK* model does not account for the presence of neutral mutations (i.e., mutations that do not change the fitness value), two different adaptations of the model incorporate this feature: the *NKP* model, where a fraction P of the fitness contributions have a value of zero, and the *NKQ* model, in which each fitness contribution can take only one of Q possible values. In the limits $P \to 0$ and $Q \to \infty$, the *NKP* and *NKQ* models correspond to the original *NK* model (10, 37).

Since its initial application to the maturation of the immune response (51), the *NK* model has been used to describe experimental protein and DNA fitness landscapes (30, 39, 88). Rugged regions in a landscape are described by high values of K, which can be estimated from the data, for example, by calculating the autocorrelation function for different values of K and comparing it to

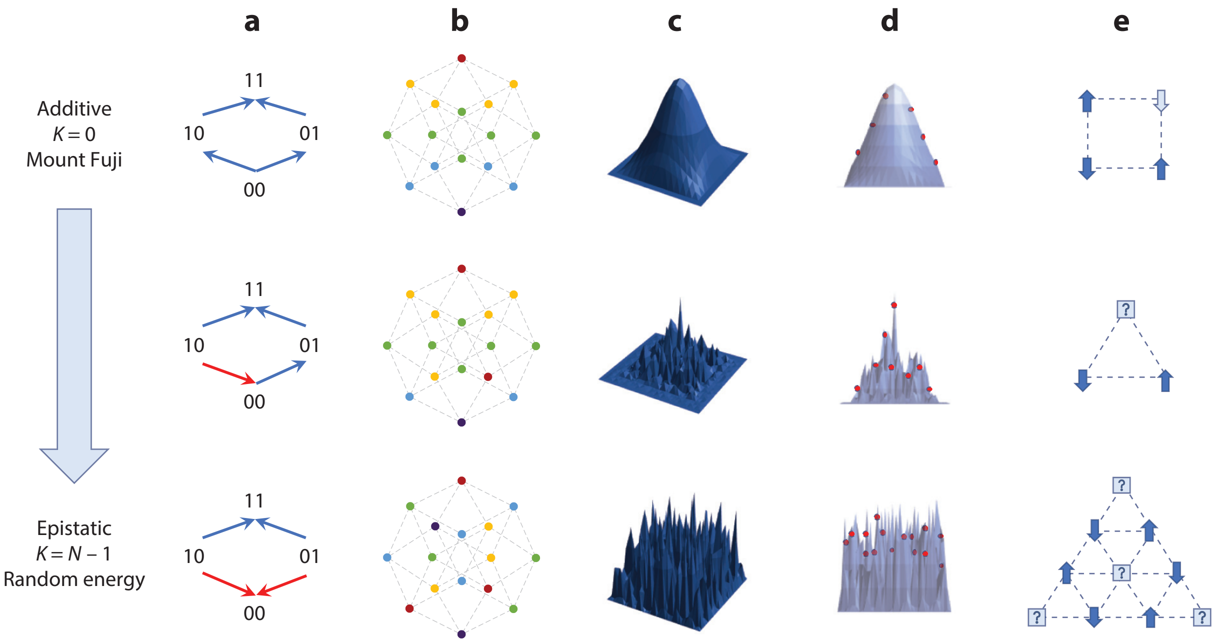

Figure 2

Epistasis and ruggedness on a fitness landscape. (*a*) For the simplest possible case ($m = 2, N = 2$), a smooth landscape can be climbed upward from 00 to 11 (peak). Sign epistasis prevents passage over one trajectory, and reciprocal sign epistasis blocks both pathways (111). Fitness increase or decline is indicated by blue or red arrows, respectively. (*b*) A similar pattern can be seen for $m = 2, N = 4$ (refer to **Figure 1*c***). (*c*) A conceptual three-dimensional depiction of fitness landscapes with varying ruggedness; horizontal axes correspond to sequence space and the vertical axis corresponds to fitness values. (*d*) Random sampling (*red dots*) can yield a better representation of smooth landscapes than of rugged ones. (*e*) Representation of frustration (or lack of) in a geometrical lattice of spins. With a smooth landscape, conditions leading to maximum fitness can be satisfied simultaneously. At high K, conditions leading to maximum fitness (or minimum energy) conflict with one another and frustrate optimization. Abbreviations: K, epistatic degree; m, alphabet size; N, number of dimensions.

the experimental system (30, 88). It is important to note that, since regions of the fitness landscape that are populated with closely related sequences of low fitness are described by $K \approx 0$, attempts to fit the *NK* model to landscapes over wide regions might result in artificially low values of K owing to averaging over dissimilar regions of the landscape. Different parameters have also been proposed to measure epistasis in fitness landscapes (e.g., number of peaks, ratio of the roughness over additive fitness, or fraction of sign epistasis). Ferretti et al. (23) recently proposed a new measure more directly related to epistasis—namely, the single-step correlation of fitness effects for mutations between neighbor genotypes, which can also be used in landscapes with missing data. A summary of calculations of this measure for different theoretical landscapes can be found in Reference 23.

Tunable ruggedness can also be introduced into the Mount Fuji model (4, 64). The rough Mount Fuji model is the addition of a Mount Fuji–type landscape and the uncorrelated house of cards model. This model can include sign epistasis, in which the effect of a single mutant is positive or negative depending on the presence of another mutation (e.g., **Figure 2*a***), provided there exists a different single mutant that is more fit than the double mutant. The ruggedness is tuned by varying the proportion of additive and random fitness components. Examples of landscapes with varying ruggedness are given in **Figure 2**.

Sign epistasis: epistasis in which one mutation has the opposite effect on fitness when in the presence of another mutation

4. CASE STUDY ON EVOLUTIONARY OPTIMIZATION: NEUTRAL VERSUS FRUSTRATED NETWORKS

Neutral network: a network of evolutionary pathways in which fitness changes are negligible

Frustration: property describing a system that cannot simultaneously satisfy constraints of maximum favorability for each variable component

Fitness peak: family of related sequences with elevated fitness

An important property of any fitness landscape is the ease with which evolution can optimize fitness. Whether this is feasible depends on the ruggedness of the landscape and specifically on whether viable evolutionary pathways (i.e., uphill climbs under SSWM) allow access to the global optimum from distant areas of sequence space. Early computational work investigating this problem studied whether viable paths could be found connecting unrelated RNA sequences that were predicted to fold into the same secondary structure. These simulations, which took advantage of the high accuracy of RNA secondary structure prediction (56), required conservation of the fold to define a viable path. These simulations revealed two related insights. First, they predicted that almost all common folds occur within any small region of sequence space (26). Second, for common folds, the large set of sequences that share a given fold would form an evolutionary network throughout sequence space (27, 46, 95, 102). The fact that this set is large is important; if the fraction of sequence space that adopts the desired fold is low, then the folded sequences represent isolated regions in the space. However, if the fraction reaches a critical percolation threshold ($\sim 1/N$), the islands become connected and the landscape as a whole exhibits a neutral network (34). A neutral network could be conceptualized as a fitness landscape topography that is full of "holes," emphasizing the fact that high-dimensional sequence space has a nonintuitively vast number of potential connections (35). These computational and theoretical considerations gave rise to the attractive hypothesis that neutral networks might characterize molecular fitness landscapes, allowing evolutionary optimization over large distances.

In contrast to this view of neutral networks, many empirical examples of epistasis are known in local sequence space, and one might expect that the extension of epistasis through the landscape (i.e., widespread ruggedness) would result in frustrated optimization during selection. This phenomenon can be mimicked in the *NK* model, which can be interpreted as a superposition of *p*-spin glass models (99) (**Figure 2*e***). In spin glasses, the Hamiltonian of the system exhibits frustration when no spin configuration can simultaneously satisfy all couplings, leading to a state of minimum energy. Since there is no single lowest-energy configuration, the energy landscape contains several metastable states separated by a distribution of energy barriers. The parameter p (number of interacting spin glasses) tunes the ruggedness of the energy landscape, much like K in the *NK* model. In the limit $p \to \infty$, it becomes impossible to satisfy all spin constraints and the system has an extremely rugged, uncorrelated potential surface, equivalent to Derrida's random energy model (18), which is an analog of the random house of cards model. Similarly, in the *NK* model, as K increases, configurations leading to the highest fitness contribution for certain positions become mutually incompatible, leading to blocked evolutionary paths over which optimization by selection is frustrated.

Ideally, experimental detection of a neutral versus frustrated network would involve mapping the topography of a complete fitness landscape. However, owing to the large size of sequence space for even small folded RNAs and the limits of sequencing throughput at the time, early work related to this question focused on construction of a viable evolutionary pathway between two nucleic acid sequences with different functions (15, 41, 94). Several examples of protein evolution to produce new or altered function were also known (e.g., 40). These efforts were surprisingly successful, suggesting that different functions could be nearby in sequence space; in other words, fitness peaks for different functions can overlap.

Nevertheless, investigating evolutionary optimization on a single fitness landscape requires identification of a very large number of functional sequences, and thus substantial progress had to await the advent of HTS. The first complete fitness landscape for short RNA sequences

($N = 21$) revealed very few viable evolutionary paths between different functional families (48). Although this approach cannot be easily extended to much longer lengths, one attempt to evolve an RNA polymerase ribozyme ($N = 168$) at a high mutation rate did not find a new optimum (71). Although this careful study was able to relate the results of the selection to the topography of the fitness landscape, it is possible that similar results in other systems are underreported in the literature. These studies hint that frustration may characterize evolutionary optimization of a particular function for RNA for a relatively fixed landscape. Given the contrast between these frustrated cases and the apparent ease of evolving certain new functions, it is tempting to speculate that optimization of a single function might have quite different evolutionary properties than evolution of a new function.

Ribozyme: RNA sequence that catalyzes a specific reaction

Aptamer: an RNA (or DNA) that binds to a specific ligand

In vitro selection (or evolution): laboratory evolution of biomolecules that selects sequences from a pool of variants based on ability to carry out a specific function

5. MEASURING MOLECULAR FITNESS LANDSCAPES WITH HIGH-THROUGHPUT TECHNIQUES

5.1. RNA and DNA: From Microarrays to High-Throughput Sequencing

When measuring fitness landscapes, functional nucleic acids present certain advantages compared to more complicated evolvable systems. In particular, an alphabet of only four nucleotides allows far higher coverage of random sequence libraries. Predominantly in silico approaches have shown some utility in predicting activity, such as in the generation of an effective anti-HIV aptamer (an RNA-based affinity reagent) (91), but such studies are relatively uncommon. On the experimental side, HTS for studying fitness landscapes can be seen as the successor high-throughput technique following microarrays, paralleling the trend in genomics applications. Approximately 10^5–10^6 sequences can be studied in reasonable copy number with a single HTS run (or microarray assay), equivalent to full coverage of sequence space with $N = 10$. Nucleic acid microarrays have been used to investigate double- and triple-mutational scans of aptamers (49), used with rational truncation to investigate the importance of structural constraints on aptamer activity (25) and combined with in silico approaches to interrogate large local evolutionary spaces in array-based directed evolution (54). A 2010 study (88) was able to use array techniques to measure DNA–protein binding over all possible 10-nucleotide sequences, showing that although the fitness landscape contained only a single conserved active motif, the landscape contained sufficient ruggedness to produce many separate local fitness optima.

But microarray approaches have been somewhat limited in their scope and adoption for multiple reasons, including their reliance on reactions or binding events producing a fluorescent signal and limitations stemming from attachment of the nucleic acid to a surface. Instead, HTS-based approaches have increasingly come to dominate RNA and DNA fitness landscape studies (7). In 2010, Pitt & Ferré-D'Amaré (77) demonstrated the ability of HTS to measure sequence enrichment during in vitro selection as an estimate of sequence fitness, generating a local landscape of approximately 10^7 mutant variants of a ligase ribozyme (catalytic RNA; **Figure 3**). The increasing scale and affordability of HTS technology has made such measurements an accessible option. Further development of HTS measurement of fitness landscapes has focused on techniques to improve either landscape coverage or measurement of fitness.

To improve landscape coverage and interrogate larger sequence spaces, the limitation is not pool size (typically 10^{14}–10^{16} molecules) but analytical capability [i.e., sequencing throughput (typically 10^6–10^8 reads)]. It is possible to overcome this limit with in vitro selection—if selection can isolate nearly all of the high-activity sequences, complete mapping of an RNA fitness landscape becomes possible for short sequences. When studying molecular fitness landscapes in vitro, the interpretation of negative information can be powerful (48). This requires a well-defined initial pool but potentially expands the analysis, as it is no longer limited by the sequencing throughput

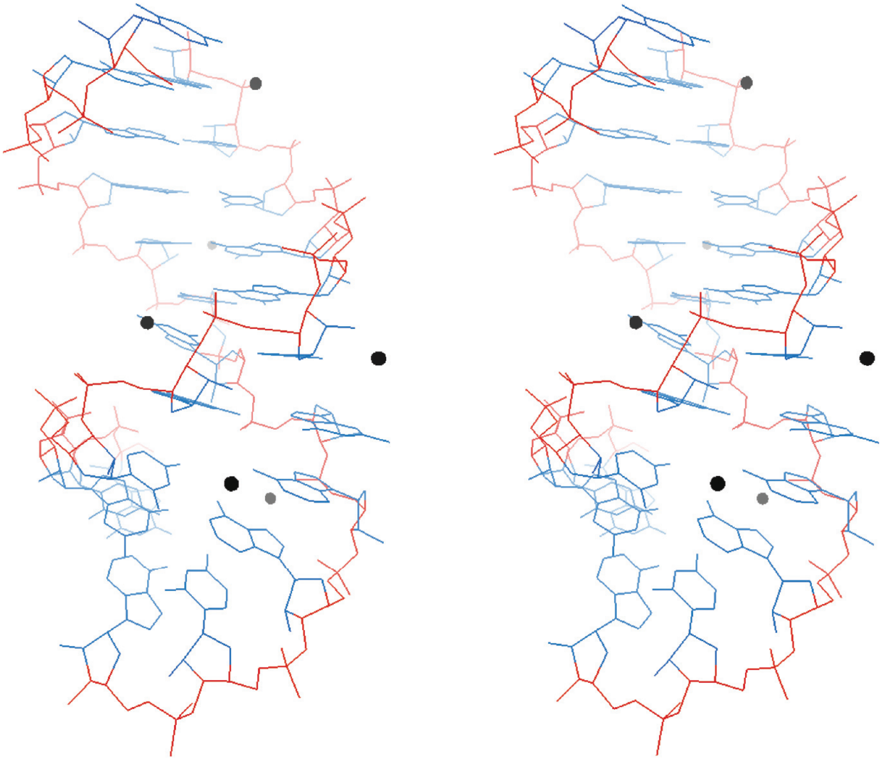

Figure 3

Stereo view of the structure of the class II ligase ribozyme (76) PDB ID: 3FTM; image created with Visual Molecular Dynamics (45).

but by the complexity of the initial pool, which is larger by several orders of magnitude. Although detailed information cannot be obtained about lost mutants, their disappearance indicates low fitness. It should be noted that epistasis and other studies should be interpreted with respect to the mutants analyzed. For example, if the mutants are not selected at random (e.g., survived a selection), epistasis values for that subpopulation would likely underestimate those for random mutants unless negative information is taken into account. At the same time, sparse random sampling can also lead to inaccurate estimation of epistasis and ruggedness (68), and the prevalence of indirect evolutionary pathways that bypass local valleys (114) could lead to underestimates of evolvability if the explored space is too small. However, depending on the hypothesis or question being investigated, in vitro selections from a large, random pool that only sparsely covers sequence space can still provide insights into general underlying trends in the larger, unmeasurable spaces (77, 79).

For in vitro selection experiments, fitness is taken to reflect chemical activity and can be estimated (or defined) in multiple ways, such as abundance at the end of selection, enrichment over a single round, or functional activity under selection conditions. Ideally, all of these should be correlated, as they are related to the true chemical activity of a given selected species. Abundance, however, can be surprisingly poorly correlated to chemical activity (48, 79), likely owing to experimental noise and biases related to sequencing (e.g., polymerase chain reaction). Thus, new approaches use HTS to perform direct activity screens (19, 47, 55). Furthermore, fitness estimates can be notably improved by considering multiple rounds of selection (79).

High-throughput techniques are also being applied to measurement of RNA and DNA specificity. While these experiments often address different scientific questions than single-function fitness landscapes, they use similar techniques and analyses. HTS techniques were used to characterize the DNA binding landscapes of over a thousand transcription factors (TFs) (2). These data enabled mapping of DNA–TF binding energy over large sequence spaces (57), again illustrating the power of applying HTS to traditional questions.

5.2. Beyond DNA and RNA: Exploring New Chemical Space with High-Throughput Sequencing

Recent forays into the chemical space of nucleic acids with altered backbones (XNAs) or modified bases raise the prospect that, with modern knowledge and techniques, parallel molecular biology could be developed for these alternative nucleic acids in a relatively short time (74). Alternative nucleic acids raise many fundamental questions about fitness landscapes, from biologically inspired issues such as the uniqueness (or lack thereof) of RNA and DNA to more abstract problems, such as the shape of the larger fitness landscape in chemical space. While chemical study of alternative nucleic acids dates back to Eschenmoser's pioneering work (22), investigations into their functional capacity began with altered bases—namely, in vitro selection on reduced alphabets. Remarkably, ribozymes could be made from alphabets of only three (85, 86) or even two letters (82). In both cases, reduction in alphabet size led to selected ribozymes with lower activity than their larger-alphabet counterparts. In contrast, artificially expanded genetic information systems (AEGISs) employ additional letters (96, 116) and have been used to identify six-letter aptamers with greater affinity than those selected containing four letters. While AEGIS currently poses some complications requiring probabilistic decoding of HTS data, HTS may still be applied to increase throughput compared to Sanger sequencing. Further advances to functionality are aided by a wider exploration of bases. For example, the incorporation of unnatural hydrophobic nucleobases [e.g., 7-(2-thienyl)imidazo[4,5-*b*]pyridine (Ds), SOMAmers, click-SELEX] result in increased binding affinity to their protein targets (36, 52, 105).

For some functions, the activity of functional DNA molecules is comparable to that of RNA molecules (97). On some occasions, the sequence of a functional RNA can be simply synthesized as DNA and retain functionality (106, 109), sometimes requiring additional evolution (69). These exceptional cases may arise if the major interactions are electrostatic or nonspecific stacking interactions. XNAs made from nonnatural backbone alterations (**Figure 4**) have been selected for binding and catalytic activity, with activities similar to those seen in natural nucleic acids (75, 103, 115). Introduction of phosphorodithioate linkages can improve aptamer binding (108), with a single modified linkage increasing affinity by ~1,000-fold in one case (1). Another aspect of fitness is the chemical and physiological stability of the molecule; for example, many backbone modifications confer resistance to ribonuclease degradation (14). Other modifications, such as 2′-fluoro and 2′-amino RNA, provide both added stability (73) and sometimes increased functionality (104). The employment of chemical modifications to improve nucleic acids has been reviewed in more detail in References 21, 65, and 87.

The application of HTS to alternative nucleic acids is not trivial because of the need for engineered polymerases to accept the template and read it out in a decodable way. Still, these challenges are being overcome by ingenious strategies (75, 96, 103). Although XNA fitness landscapes are largely unstudied at the moment, it seems inevitable that some may demonstrate different or higher fitness peaks. Whether these changes will lead to new evolutionary properties is currently a fascinating unknown.

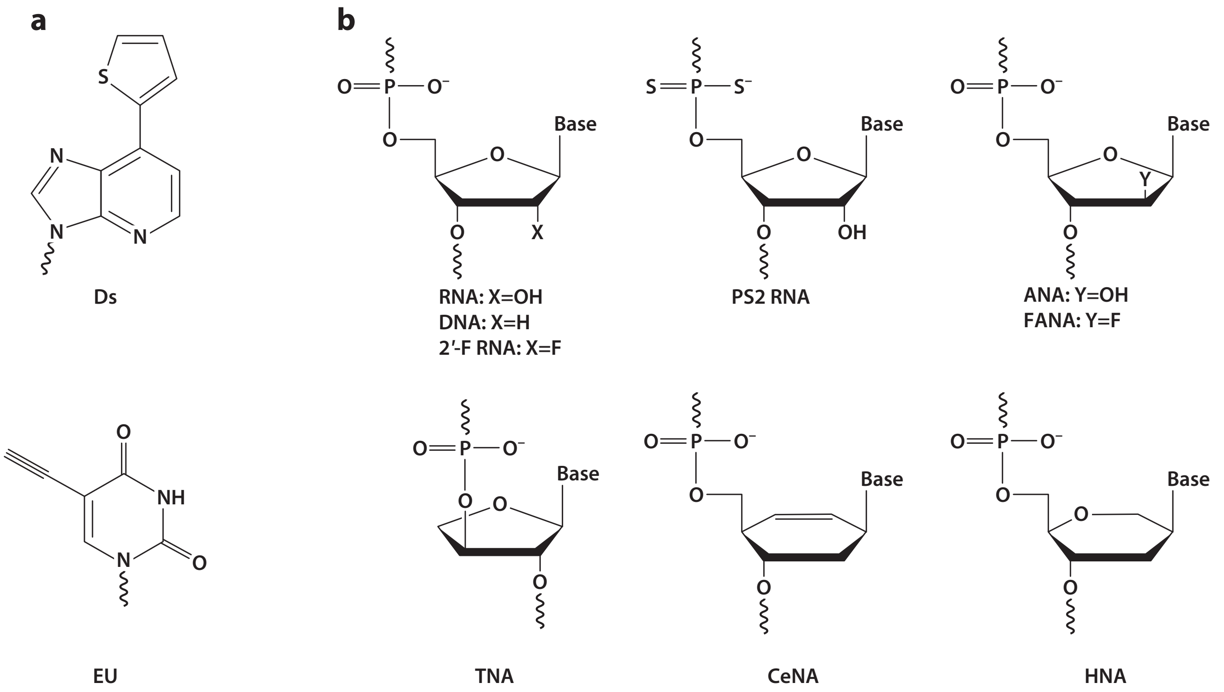

Figure 4

Expanded chemical space of functional nucleic acids. (*a*) The modified bases Ds [7-(2-thienyl)imidazo[4,5-b]pyridine] and EU (C5-ethynyl-uracil), which is utilized in click-SELEX. (*b*) Chemical structures for RNA, DNA, 2′-F RNA (2′-fluoro RNA), ANA (arabino nucleic acid), FANA (2′-fluoro ANA), PS2 RNA (phosphorodithioate RNA), TNA (threose nucleic acid), CeNA (cyclohexenyl nucleic acid), and HNA (1,5-anhydrohexitol nucleic acid).

6. FITNESS LANDSCAPES OF ORGANISMS: RNA, PROTEINS, AND GENOMES

Complete coverage of sequence space for an organismal genome—or even a single gene—is intractable owing to the size of sequence space involved. However, local sampling around functional proteins (or random sampling of genomic mutants) still provides a rich source of data about the local landscape of the protein or the organism as a whole. Some examples of ways to represent HTS data are shown in **Figure 5**. Fitness landscape studies on sequences in vivo access fewer individuals ($\sim 10^{12}$ cells in 1 L) compared to in vitro studies. While this limits the diversity of the starting pool, it does not directly affect the number of mutants that can be assayed, since sequencing throughput is still limiting.

The in vivo fitness landscapes of small functional (noncoding) RNAs [tRNA (transfer RNA) and snoRNAs (small nucleolar RNAs)] in yeast have been investigated using HTS to study all single and double mutants. Because these cellular RNAs have smaller sequence spaces than proteins, such experiments can be done at higher mutational coverage, providing a good system for exploring in vivo fitness landscapes. In these cases, coverage of the local area around the wild-type sequence indicates that epistatic effects of mutation tend to be negative, with loss of fitness often corresponding to predicted disruption of RNA folding (59, 80). As more RNA fitness landscapes are examined, it will be interesting to compare landscape characteristics of highly evolved biological RNAs versus RNAs evolved in vitro to understand how >3.5 billion years of natural selection

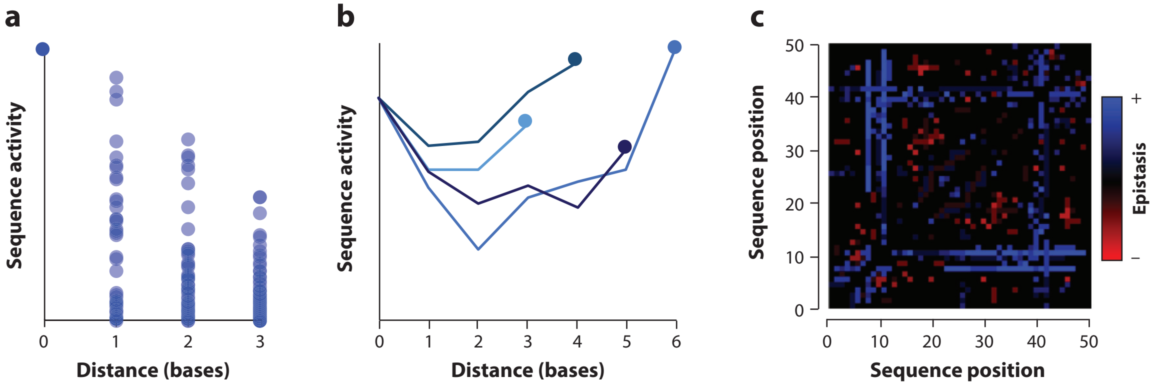

Figure 5

Representing high-throughput sequencing data of fitness landscapes. (*a*) A fitness peak with sequence space collapsed onto one dimension representing the edit distance (i.e., number of mutations) from the optimum sequence. Panel *a* adapted from References 48, 77, and 88. (*b*) Evolutionary pathways between one local optimum and other nearby local optima, with sequence space collapsed as in panel *a*. This representation illustrates fitness valleys and ruggedness. Panel *b* adapted from References 48 and 88. (*c*) Heat map representing combinations of mutants, revealing epistatic interactions along the length of a sequence. Panel *c* adapted from References 12 and 80.

has shaped the landscape itself. Furthermore, the introduction of modified bases into cells (62) suggests the intriguing possibility of measuring fitness landscapes of alternative nucleic acids in vivo.

The study of protein fitness landscapes, which began with mutational analysis (e.g., alanine scanning) and combinatorial studies of selected mutants, has been greatly impacted by HTS. Both m and N are substantially greater for proteins than RNA (e.g., the number of single-mutant variants to be tested would be ∼6,000 for a typical single-domain protein of length of ∼300 amino acids, compared to ∼150 single-mutant variants tested for a typical ribozyme with a length of 50 nucleotides). The jump from Sanger sequencing to HTS has increased the number of mutants that can be analyzed by at least 4 orders of magnitude.

In an HTS technique known as deep mutational scanning (DMS), the activity of a mutant library is linked to organismal (cell or virus) fitness (5) [e.g., by cell sorting or simply by reproduction and survival for influenza variants (72)]; DMS has been further reviewed (29, 100). The survival of cells (or viruses) harboring the mutant library is measured by HTS, allowing assay of the fitness effect of 10^5–10^6 protein variants. DMS has proven effective for creating high-coverage, highly local fitness landscapes centered around a wild-type protein and can identify sites of conserved function (28). The local fitness landscape of the green fluorescent protein, measured over thousands of derivative genotypes, was found to be quite narrow, with the majority of single mutants showing reduced fluorescence (92). In contrast, DMS of a complete 9–amino acid region of Hsp90 showed that the distribution of fitness was bimodal, with one mode consisting of nearly neutral mutations and the other of deleterious mutations (43). On a practical side, DMS results within yeast were used to optimize protein engineering, resulting in a new protein (with five point mutations) with a 25-fold increase in binding affinity to the influenza virus hemagglutinin (112).

DMS is well poised to measure local epistasis of a protein, since the fitness effect of many combinations of mutations can be measured. Even so, analysis of epistasis on in vivo protein landscapes is generally limited to a small number of peptide sites, a limited library of amino acid substitutions, or one specific set of evolutionary paths (17). Weinreich et al. (110) compiled a comprehensive review of these studies, showing that in these limited-landscape cases, in vivo protein epistasis tends to be primarily dominated by low-order epistatic effects of only a few loci, although higher-order epistasis was notable in some cases. A local fitness landscape for four positions in protein GB1 revealed a very interesting feature: Although many direct evolutionary pathways were blocked by reciprocal sign epistasis, these evolutionary dead ends could be avoided by following indirect paths in the sequence space (114). Limited epistasis and evolutionary detours suggest short neutral pathways; whether these could combine over larger sequence space to form a neutral network is still unknown. However, sequencing technology continues to improve and may allow study of this question to be taken further in the future.

Although the theoretical models described earlier are highly simplified, one may ask whether empirical fitness landscapes can be fit to them. One 2013 meta-analysis found general trends in ruggedness and epistasis across a number of such studies, with many showing reasonable agreement with patterns expected from a rough Mount Fuji model (101). Efforts to connect empirical data to these models are important for gaining an intuitive grasp of the topography of fitness landscapes. It remains an open question whether these models can also describe effects over organismal fitness landscapes of a larger scale, multiple peaks, or covering evolutionary sites on multiple genes.

7. ENVIRONMENT AND THE FITNESS LANDSCAPE

It is nearly impossible to overstate the importance of the environment in determining the topography of a fitness landscape (**Figure 6**). At the microscopic level, molecular fitness depends on the temperature, water activity, pH, phase, cosolutes, and nearly any other environmental variable. These effects modulate both basic properties [e.g., RNA stability (33)] as well as sophisticated functions [e.g., ribozyme activity (8, 31, 93)]. At the macroscopic level, genetic and environmental effects on traits cannot be simply deconvolved, as the heritability of any trait depends on the environment and genetic background in which it is measured. Even without environmental perturbations, the fitness landscape of a metabolizing organism is a continuously dynamic object, as organisms modify their environment, which changes the fitness landscape. Perhaps the most well-known example of this comes from the multi-decade experimental evolution of *Escherichia coli*, in which changes to the genetic background (potentiating mutations) enabled evolution of the ability to metabolize citrate (11). The efforts may also be driven by the potential for biomedical applications as well; for example, DMS of a kinase involved in antibiotic resistance demonstrates a fitness landscape that varies significantly over changes in both antibiotic concentration and

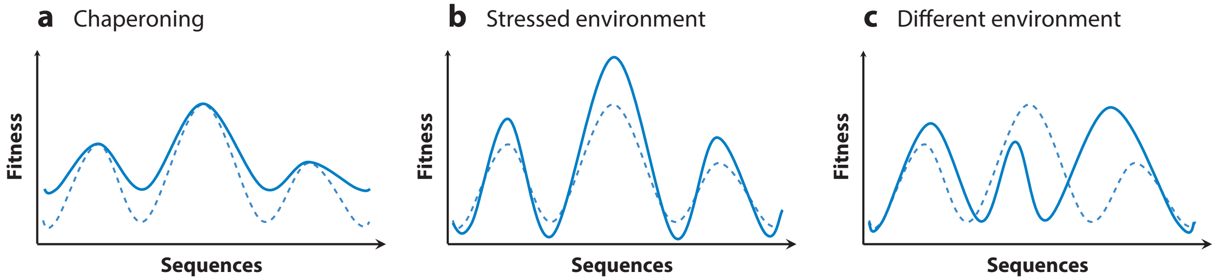

Figure 6

The fitness landscape depends strongly on the environment. For molecular fitness landscapes, environments might confer (*a*) stabilization of weakly folded structures (chaperoning), (*b*) exaggeration of fitness differences under stressed conditions, or (*c*) completely different structure in a new environment. The illustrations indicate the fitness landscape in one environment (*dotted line*) and in a new environment (*solid line*).

structure (63). Systematic study of the effect of the environment on the fitness landscape using HTS represents a major goal for this field.

Riboswitch: *cis*-acting RNA element containing an aptamer, in which ligand binding alters transcription or translation

The importance of the environmental context can be seen even in relatively simple molecular fitness landscapes for RNA. While most studies of functional RNA occur in vitro, it is clear that in vivo conditions may differ, sometimes greatly. For example, aptamer-based biosensors evolved in vitro show significantly lower performance in blood than in buffer (6). Crowded and confined conditions can modify the structure and function of nucleic acids and proteins (16, 32, 83, 89, 90). High levels of molecular crowding have been shown to stabilize mutations in ribozymes (58), change the binding mechanism of a ligand to a riboswitch (84), and create a chaperoning effect to assist in aptamer folding (90). Ribozymes can also modify their environment [e.g., through cooperation (107)], presenting an attractive future target for mapping more complex fitness landscapes.

To study the effect of the environment on organismal landscapes, one common method is to expose the population to a new environment and observe the resulting evolution. In general, organismal fitness drops after environmental changes but largely recovers through subsequent evolution and delayed adaptation at the genetic level (24, 44). For example, changes to the fitness landscape of Hsp90 in *Saccharomyces cerevisiae* were observed in elevated salinity, with previously adaptive mutations becoming deleterious in the new environment (42), and the accessible evolutionary pathways in an esterase were shown to change at different growth temperatures (67). Interestingly, variation in hosts may alter the topology of a viral fitness landscape, which may drive virus specialization (13). However, whether the fitness landscape of a gene varies in different environments seems to depend on the details of the system. In contrast to cellular proteins, where a gene's fitness contribution often does vary with environment, studies of tRNA indicate that mutations influence the gene's fitness contribution by a fixed proportion independent of the environment, for four growth environments tested (60). Further work in the yeast tRNA system also indicates that epistatic effects between loci can vary significantly for the same gene between different organisms (20). If a mutation has multiple conflicting effects on fitness (antagonistic pleiotropy), adaptation to a new environment might be limited. Landscape analysis of the yeast genome shows that many gene variants display some degree of antagonistic pleiotropy in specific growth conditions (81). The environmental landscape for a single sequence can also be measured, as was done for a riboswitch in nearly 20,000 different environmental conditions (9). Measurement of such environmental landscapes in conjunction with fitness landscapes is a challenging but essential goal for which high-throughput techniques are essential.

8. OUTLOOK

High-throughput sequencing has transformed the study of fitness landscapes, expanding the focus from theoretical models to empirical mapping. Increased sequencing throughput is more than a quantitative extension, as it allows exploration of fundamentally new areas of science, from evolutionary networks to environmental landscapes. To maximize the knowledge return from this exciting growth of data, perhaps two aspects should be kept in mind. First, attention should be paid to building intuition and understanding, such as by analyzing the fit of data to idealized model landscapes. Second, while raw HTS data can be submitted to databases such as the National Center for Biotechnology Information Sequence Read Archive, a dedicated resource for submitting and viewing fitness landscape data could facilitate meta-analysis, standardization, and contributions from a greater community of researchers. Regardless, HTS-enabled mapping of fitness landscapes brings the tantalizing prospect of predicting evolution closer yet to reality.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Funding from the Simons Foundation (grant 290356), NASA (grant NNX16AJ32G), and the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office is gratefully acknowledged. The content of the information does not necessarily reflect the position or the policy of the US government, and no official endorsement should be inferred.

LITERATURE CITED

1. Abeydeera ND, Egli M, Cox N, Mercier K, Conde JN, et al. 2016. Evoking picomolar binding in RNA by a single phosphorodithioate linkage. *Nucleic Acids Res.* 44:8052–64
2. Aguilar-Rodríguez J, Payne JL, Wagner A. 2017. A thousand empirical adaptive landscapes and their navigability. *Nat. Ecol. Evol.* 1:0045
3. Aita T, Husimi Y. 1996. Fitness spectrum among random mutants on Mt. Fuji-type fitness landscape. *J. Theor. Biol.* 182:469–85
4. Aita T, Uchiyama H, Inaoka T, Nakajima M, Kokubo T, Husimi Y. 2000. Analysis of a local fitness landscape with a model of the rough Mt. Fuji-type landscape: application to prolyl endopeptidase and thermolysin. *Biopolymers* 54:64–79
5. Araya CL, Fowler DM. 2011. Deep mutational scanning: assessing protein function on a massive scale. *Trends Biotechnol.* 29:435–42
6. Arroyo-Currás N, Dauphin-Ducharme P, Ortega G, Ploense KL, Kippin TE, Plaxco KW. 2018. Subsecond-resolved molecular measurements in the living body using chronoamperometrically interrogated aptamer-based sensors. *ACS Sens.* 3:360–66
7. Athavale SS, Spicer B, Chen IA. 2014. Experimental fitness landscapes to understand the molecular evolution of RNA-based life. *Curr. Opin. Chem. Biol.* 22:35–39
8. Attwater J, Wochner A, Holliger P. 2013. In-ice evolution of RNA polymerase ribozyme activity. *Nat. Chem.* 5:1011
9. Baird NJ, Inglese J, Ferré-D'Amaré AR. 2015. Rapid RNA–ligand interaction analysis through high-information content conformational and stability landscapes. *Nat. Commun.* 6:8898–98
10. Barnett L. 1998. Ruggedness and neutrality—the NKp family of fitness landscapes. In *Proceedings of the Sixth International Conference on Artificial Life*, pp. 18–27. Cambridge, MA: MIT Press
11. Blount ZD, Borland CZ, Lenski RE. 2008. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *PNAS* 105:7899–906
12. Buenrostro JD, Araya CL, Chircus LM, Layton CJ, Chang HY, et al. 2014. Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat. Biotechnol.* 32:562–68
13. Cervera H, Lalić J, Elena SF. 2016. Effect of host species on the topography of fitness landscape for a plant RNA virus. *J. Virol.* 90:10160–69
14. Culbertson MC, Temburnikar KW, Sau SP, Liao JY, Bala S, Chaput JC. 2016. Evaluating TNA stability under simulated physiological conditions. *Bioorg. Med. Chem. Lett.* 26:2418–21
15. Curtis EA, Bartel DP. 2005. New catalytic structures from an existing ribozyme. *Nat. Struct. Mol. Biol.* 12:994–1000
16. Daher M, Widom JR, Tay W, Walter NG. 2018. Soft interactions with model crowders and non-canonical interactions with cellular proteins stabilize RNA folding. *J. Mol. Biol.* 430:509–23

17. de Visser JA, Krug J. 2014. Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.* 15:480–90

18. Derrida B. 1981. Random-energy model: an exactly solvable model of disordered systems. *Phys. Rev. B* 24:2613–26

19. Dhamodharan V, Kobori S, Yokobayashi Y. 2017. Large scale mutational and kinetic analysis of a self-hydrolyzing deoxyribozyme. *ACS Chem. Biol.* 12:2940–45

20. Domingo J, Diss G, Lehner B. 2018. Pairwise and higher-order genetic interactions during the evolution of a tRNA. *Nature* 558:117–21

21. Dunn MR, Jimenez RM, Chaput JC. 2017. Analysis of aptamer discovery and technology. *Nat. Rev. Chem.* 1:0076

22. Eschenmoser A. 1999. Chemical etiology of nucleic acid structure. *Science* 284:2118–24

23. Ferretti L, Schmiegelt B, Weinreich D, Yamauchi A, Kobayashi Y, et al. 2016. Measuring epistasis in fitness landscapes: the correlation of fitness effects of mutations. *J. Theor. Biol.* 396:132–43

24. Filteau M, Hamel V, Pouliot M-C, Gagnon-Arsenault I, Dubé AK, Landry CR. 2015. Evolutionary rescue by compensatory mutations is constrained by genomic and environmental backgrounds. *Mol. Syst. Biol.* 11:832

25. Fischer NO, Tok JB-H, Tarasow TM. 2008. Massively parallel interrogation of aptamer sequence, structure and function. *PLOS ONE* 3:e2720

26. Fontana W, Konings DA, Stadler PF, Schuster P. 1993. Statistics of RNA secondary structures. *Biopolymers* 33:1389–404

27. Fontana W, Schuster P. 1998. Continuity in evolution: on the nature of transitions. *Science* 280:1451–55

28. Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, et al. 2010. High-resolution mapping of protein sequence-function relationships. *Nat. Methods* 7:741–46

29. Fowler DM, Fields S. 2014. Deep mutational scanning: a new style of protein science. *Nat. Methods* 11:801–7

30. Franke J, Klozer A, de Visser JA, Krug J. 2011. Evolutionary accessibility of mutational pathways. *PLOS Comput. Biol.* 7:e1002134

31. Frommer J, Appel B, Müller S. 2015. Ribozymes that can be regulated by external stimuli. *Curr. Opin. Biotechnol.* 31:35–41

32. Gao M, Held C, Patra S, Arns L, Sadowski G, Winter R. 2017. Crowders and cosolvents—major contributors to the cellular milieu and efficient means to counteract environmental stresses. *Chem. Phys. Chem.* 18:2951–72

33. Gao M, Arns L, Winter R. 2017. Modulation of the thermodynamic signatures of an RNA thermometer by osmolytes and salts. *Angew. Chem. Int. Ed.* 56:2302–6

34. Gavrilets S. 2004. *Fitness Landscapes and the Origin of Species (MPB-41)*. Princeton, NJ: Princeton Univ. Press

35. Gavrilets S. 2014. Models of speciation: Where are we now? *J. Hered.* 105(Suppl. 1):743–55

36. Gawande BN, Rohloff JC, Carter JD, Von Carlowitz I, Zhang C, et al. 2017. Selection of DNA aptamers with two modified bases. *PNAS* 114:2898–903

37. Geard N, Wiles J, Hallinan J, Tonkes B, Skellett B. 2002. A comparison of neutral landscapes—NK, NKp and NKq. In *Proceedings of the 2002 Congress on Evolutionary Computation*. Piscataway, NJ: IEEE

38. Gillespie JH. 1983. Some properties of finite populations experiencing strong selection and weak mutation. *Am. Nat.* 121:691–708

39. Hayashi Y, Aita T, Toyota H, Husimi Y, Urabe I, Yomo T. 2006. Experimental rugged fitness landscape in protein sequence space. *PLOS ONE* 1:e96

40. Hayashi Y, Sakata H, Makino Y, Urabe I, Yomo T. 2003. Can an arbitrary sequence evolve towards acquiring a biological function? *J. Mol. Evol.* 56:162–68

41. Held DM, Greathouse ST, Agrawal A, Burke DH. 2003. Evolutionary landscapes for the acquisition of new ligand recognition by RNA aptamers. *J. Mol. Evol.* 57:299–308

42. Hietpas RT, Bank C, Jensen JD, Bolon DNA. 2013. Shifting fitness landscapes in response to altered environments. *Evolution* 67:3512–22

17. Accessible review on fitness landscapes and their connection to evolutionary questions.

20. Comprehensive work measuring epistatic effects over a local tRNA fitness landscape in multiple genetic backgrounds.

29. Describes DMS, an important new HTS tool for studying local protein fitness landscapes.

43. Hietpas RT, Jensen JD, Bolon DN. 2011. Experimental illumination of a fitness landscape. *PNAS* 108:7896–901

44. Ho W-C, Zhang J. 2018. Evolutionary adaptations to new environments generally reverse plastic phenotypic changes. *Nat. Commun.* 9:350

45. Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38

46. Huynen MA. 1996. Exploring phenotype space through neutral evolution. *J. Mol. Evol.* 43:165–69

47. Jalali-Yazdi F, Lai LH, Takahashi TT, Roberts RW. 2016. High-throughput measurement of binding kinetics by mRNA display and next-generation sequencing. *Angew. Chem. Int. Ed.* 55:4007–10

48. Jiménez JI, Xulvi-Brunet R, Campbell GW, Turk-MacLeod R, Chen IA. 2013. Comprehensive experimental fitness landscape and evolutionary network for small RNA. *PNAS* 110:14984–89

48. Complete mapping of functional RNA fitness landscape through HTS and in vitro selection (4^{21} sequences).

49. Katilius E, Flores C, Woodbury NW. 2007. Exploring the sequence space of a DNA aptamer using microarrays. *Nucleic Acids Res.* 35:7626–35

50. Kauffman S, Levin S. 1987. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* 128:11–45

51. Kauffman SA, Weinberger ED. 1989. The NK model of rugged fitness landscapes and its application to maturation of the immune response. *J. Theor. Biol.* 141:211–45

52. Kimoto M, Yamashige R, Matsunaga K-i, Yokoyama S, Hirao I. 2013. Generation of high-affinity DNA aptamers using an expanded genetic alphabet. *Nat. Biotechnol.* 31:453–57

53. Kingman JFC. 1978. A simple model for the balance between selection and mutation. *J. Appl. Probability* 15:1–12

54. Knight CG, Platt M, Rowe W, Wedge DC, Khan F, et al. 2009. Array-based evolution of DNA aptamers allows modelling of an explicit sequence-fitness landscape. *Nucleic Acids Res.* 37:e6

55. Kobori S, Yokobayashi Y. 2016. High-throughput mutational analysis of a twister ribozyme. *Angew. Chem.* 55:10354–57

56. Kun Á, Szathmáry E. 2015. Fitness landscapes of functional RNAs. *Life* 5:1497–517

57. Le DD, Shimko TC, Aditham AK, Keys AM, Longwell SA, et al. 2018. Comprehensive, high-resolution binding energy landscapes reveal context dependencies of transcription factor binding. *PNAS* 15:E3702–11

58. Lee H-T, Kilburn D, Behrouzi R, Briber RM, Woodson SA. 2015. Molecular crowding overcomes the destabilizing effects of mutations in a bacterial ribozyme. *Nucleic Acids Res.* 43:1170–76

59. Li C, Qian W, Maclean CJ, Zhang J. 2016. The fitness landscape of a tRNA gene. *Science* 352:837–40

60. Li C, Zhang J. 2018. Multi-environment fitness landscapes of a tRNA gene. *Nat. Ecol. Evol.* 2:1025–32

61. Luksza M, Lassig M. 2014. A predictive fitness model for influenza. *Nature* 507:57–61

62. Malyshev DA, Dhami K, Lavergne T, Chen T, Dai N, et al. 2014. A semi-synthetic organism with an expanded genetic alphabet. *Nature* 509:385–88

63. Melnikov A, Rogov P, Wang L, Gnirke A, Mikkelsen TS. 2014. Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Res.* 42:e112

64. Neidhart J, Szendro IG, Krug J. 2014. Adaptation in tunably rugged fitness landscapes: the rough Mount Fuji model. *Genetics* 198:699–721

65. Ni S, Yao H, Wang L, Lu J, Jiang F, et al. 2017. Chemical modifications of nucleic acid aptamers for therapeutic purposes. *Int. J. Mol. Sci.* 18:1683

66. Obolski U, Ram Y, Hadany L. 2018. Key issues review: evolution on rugged adaptive landscapes. *Rep. Prog. Phys.* 81:012602

67. Ota N, Kurahashi R, Sano S, Takano K. 2018. The direction of protein evolution is destined by the stability. *Biochimie* 150:100–9

68. Otwinowski J, Plotkin JB. 2014. Inferring fitness landscapes by regression produces biased estimates of epistasis. *PNAS* 111:E2301–9

69. Paul N, Springsteen G, Joyce GF. 2006. Conversion of a ribozyme to a deoxyribozyme through in vitro evolution. *Chem. Biol.* 13:329–38

70. Perelson AS, Macken CA. 1995. Protein evolution on partially correlated landscapes. *PNAS* 92:9657–61

71. Petrie KL, Joyce GF. 2014. Limits of neutral drift: lessons from the in vitro evolution of two ribozymes. *J. Mol. Evol.* 79:75–90

72. Phillips AM, Gonzalez LO, Nekongo EE, Ponomarenko AI, McHugh SM, et al. 2017. Host proteostasis modulates influenza evolution. *eLife* 6:e28652
73. Pieken WA, Olsen DB, Benseler F, Aurup H, Eckstein F. 1991. Kinetic characterization of ribonuclease-resistant 2′-modified hammerhead ribozymes. *Science* 253:314–17
74. Pinheiro VB, Holliger P. 2014. Towards XNA nanotechnology: new materials from synthetic genetic polymers. *Trends Biotechnol*. 32:321–28
75. Pinheiro VB, Taylor AI, Cozens C, Abramov M, Renders M, et al. 2012. Synthetic genetic polymers capable of heredity and evolution. *Science* 336:341–44
76. Pitt JN, Ferré-D'Amaré AR. 2009. Structure-guided engineering of the regioselectivity of RNA ligase ribozymes. *J. Am. Chem. Soc.* 131:3532–40
77. **Pitt JN, Ferré-D'Amaré AR. 2010. Rapid construction of empirical RNA fitness landscapes. *Science* 330:376–79**
78. Poelwijk FJ, Tanase-Nicola S, Kiviet DJ, Tans SJ. 2011. Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *J. Theor. Biol.* 272:141–44
79. Pressman A, Moretti JE, Campbell GW, Muller UF, Chen IA. 2017. Analysis of in vitro evolution reveals the underlying distribution of catalytic activity among random sequences. *Nucleic Acids Res*. 45:10922
80. Puchta O, Cseke B, Czaja H, Tollervey D, Sanguinetti G, Kudla G. 2016. Network of epistatic interactions within a yeast snoRNA. *Science* 352:840–44
81. Qian W, Ma D, Xiao C, Wang Z, Zhang J. 2012. The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. *Cell Rep*. 2:1399–410
82. Reader JS, Joyce GF. 2002. A ribozyme composed of only two different nucleotides. *Nature* 420:841–44
83. Rivas G, Minton AP. 2016. Macromolecular crowding in vitro, in vivo, and in between. *Trends Biochem. Sci.* 41:970–81
84. Rode AB, Endoh T, Sugimoto N. 2018. Crowding shifts the FMN recognition mechanism of riboswitch aptamer from conformational selection to induced fit. *Angew. Chem. Int. Ed.* 57:6868–72
85. Rogers J, Joyce GF. 1999. A ribozyme that lacks cytidine. *Nature* 402:323–25
86. Rogers J, Joyce GF. 2001. The effect of cytidine on the structure and function of an RNA ligase ribozyme. *RNA* 7:395–404
87. Röthlisberger P, Hollenstein M. 2018. Aptamer chemistry. *Adv. Drug Deliv. Rev.* 134:3–21
88. **Rowe W, Platt M, Wedge DC, Day PJ, Kell DB, Knowles J. 2010. Analysis of a complete DNA—protein affinity landscape. *J. R. Soc. Interface* 7:397–408**
89. Saha R, Pohorille A, Chen IA. 2015. Molecular crowding and early evolution. *Orig. Life Evol. Biosph.* 44:319–24
90. Saha R, Verbanic S, Chen IA. 2018. Lipid vesicles chaperone an encapsulated RNA aptamer. *Nat. Commun.* 9:2313
91. Sanchez-Luque FJ, Stich M, Manrubia S, Briones C, Berzal-Herranz A. 2014. Efficient HIV-1 inhibition by a 16 nt-long RNA aptamer designed by combining in vitro selection and in silico optimisation strategies. *Sci. Rep.* 4:6242
92. Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, et al. 2016. Local fitness landscape of the green fluorescent protein. *Nature* 533:397–401
93. Schuabb C, Kumar N, Pataraia S, Marx D, Winter R. 2017. Pressure modulates the self-cleavage step of the hairpin ribozyme. *Nat. Commun.* 8:14661
94. Schultes E, Bartel DP. 2000. One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science* 289:448–452
95. **Schuster P, Fontana W, Stadler PF, Hofacker IL. 1994. From sequences to shapes and back: a case study in RNA secondary structures. *Proc. Biol. Sci.* 255:279–84**
96. Sefah K, Yang Z, Bradley KM, Hoshika S, Jiménez E, et al. 2014. In vitro selection with artificial expanded genetic information systems. *PNAS* 111:1449–54
97. Silverman SK. 2016. Catalytic DNA: scope, applications, and biochemistry of deoxyribozymes. *Trends Biochem. Sci.* 41:595–609

77. First work utilizing HTS to obtain local fitness landscape information about a ribozyme.

88. Early microarray work measuring the protein binding landscape for all 10-mer DNA variants.

95. In silico prediction of a neutral network for RNA secondary structure.

98. Smith JM. 1970. Natural selection and the concept of a protein space. *Nature* 225:563–64

98. Classic paper setting forth the modern understanding of a fitness landscape.

99. Stadler PF, Happel R. 1999. Random field models for fitness landscapes. *J. Math. Biol.* 38:435–78

100. Starita LM, Fields S. 2015. Deep mutational scanning: a highly parallel method to measure the effects of mutation on protein function. *Cold Spring Harb. Protoc.* 2015:711–14

101. Szendro IG, Schenk MF, Franke J, Krug J, de Visser JAGM. 2013. Quantitative analyses of empirical fitness landscapes. *J. Stat. Mech. Theory Exp.* 2013:P01005

102. Tacker M, Fontana W, Stadler P, Schuster P. 1994. Statistics of RNA melting kinetics. *Eur. Biophys. J.* 23:29–38

103. Taylor AI, Pinheiro VB, Smola MJ, Morgunov AS, Peak-Chew S, et al. 2015. Catalysts from synthetic genetic polymers. *Nature* 69:208–15

104. Thirunavukarasu D, Chen T, Liu Z, Hongdilokkul N, Romesberg FE. 2017. Selection of 2′-fluoro-modified aptamers with optimized properties. *J. Am. Chem. Soc.* 139:2892–95

105. Tolle F, Brändle GM, Matzner D, Mayer G. 2015. A versatile approach towards nucleobase-modified aptamers. *Angew. Chem. Int. Ed.* 54:10971–74

106. Travascio P, Bennet AJ, Wang DY, Sen D. 1999. A ribozyme and a catalytic DNA with peroxidase activity: active sites versus cofactor-binding sites. *Chem. Biol.* 6:779–87

107. Vaidya N, Manapat ML, Chen IA, Xulvi-Brunet R, Hayden EJ, Lehman N. 2012. Spontaneous network formation among cooperative RNA replicators. *Nature* 491:72

108. Volk DE, Yang X, Fennewald SM, King DJ, Bassett SE, et al. 2002. Solution structure and design of dithiophosphate backbone aptamers targeting transcription factor NF-κB. *Bioorg. Chem.* 30:396–419

109. Walsh R, DeRosa MC. 2009. Retention of function in the DNA homolog of the RNA dopamine aptamer. *Biochem. Biophys. Res. Commun.* 388:732–35

110. Weinreich DM, Lan Y, Jaffe J, Heckendorn RB. 2018. The influence of higher-order epistasis on biological fitness landscape topography. *J. Stat. Phys.* 172:208–25

111. Weinreich DM, Watson RA, Chao L. 2005. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* 59:1165–74

111. Explains the importance of sign and reciprocal sign epistasis for landscape topography and viable evolutionary pathways.

112. Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, et al. 2012. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat. Biotechnol.* 30:543–48

113. Wright S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In *Proceedings of the Sixth International Congress of Genetics*, Vol. 1, pp. 356–66. Menasha, WI: Brooklyn Botan. Gard.

114. Wu NC, Dai L, Olson CA, Lloyd-Smith JO, Sun R. 2016. Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife* 5:e16965

114. Demonstrates the importance of sampling over broad areas when drawing inferences about the fitness landscape.

115. Yu H, Zhang S, Chaput JC. 2012. Darwinian evolution of an alternative genetic system provides support for TNA as an RNA progenitor. *Nat. Chem.* 4:183–87

116. Zhang L, Yang Z, Sefah K, Bradley KM, Hoshika S, et al. 2015. Evolution of functional six-nucleotide DNA. *J. Am. Chem. Soc.* 137:6734–37