Title

Molecular Fitness Landscapes from High-Coverage Sequence Profiling.

Journal

Annual review of biophysics, 48(1)

ISSN

1936-122X

Authors

Blanco, Celia
Janzen, Evan
Pressman, Abe
et al.

Publication Date

2019-05-01

DOI

10.1146/annurev-biophys-052118-115333

Peer reviewed

Molecular Fitness Landscapes from High Coverage Sequence Profiling

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

*equal contribution

1. Department of Chemistry and Biochemistry 9510, University of California, Santa Barbara, CA 93106

2. Program in Chemical Engineering, University of California, Santa Barbara, CA 93106

3. Program in Biomolecular Sciences and Engineering, University of California, Santa Barbara, CA 93106

Email: blanco@ucsb.edu, evanjanzen@ucsb.edu, rsaha@ucsb.edu, ichen@ucsb.edu, abe_pressman@ucsb.edu

ORCID numbers:

Celia Blanco: 0000-0003-1536-1493

Evan Janzen: 0000-0002-1646-3363

Abe Pressman: 0000-0003-0849-620X

Ranajay Saha: 0000-0001-6001-7363

Corresponding author:

Irene A. Chen (ichen@ucsb.edu)

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.

Introduction

Predicting evolution is a key challenge in biological science which not only tests our basic understanding but also has real-world ramifications. For example, prediction of influenza virus evolution (1) 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 (2; 3). 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) (4). 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 which is daunting in scope but the source of Darwin's "endless forms most beautiful".

1. 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 some intuition, one may draw 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 be typically $<10^{17}$ 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 ~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 (5).

2. Simple models of fitness landscapes

Experimental investigation of fitness landscapes is difficult due 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 (6; 7), we introduce here two simple and influential models (Mt. Fuji and *NK*) and related models (Rough Mt. Fuji and House of Cards), to develop some intuition for possible topographies and their possible mechanisms of origin.

The simplest theoretical model is the 'Mt. Fuji' landscape (8), named after Japan's highest mountain because it is a smooth, single-peak landscape. Mt. 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 Mt. 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 are additive (the effect of each site does not depend on the others, i.e., there is no epistasis). The absence of local optima on Mt. 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 (9)), evolution on Mt. Fuji-type landscapes results in the optimal sequence.

Most empirical landscapes exhibit certain epistatic interactions that the Mt. Fuji model cannot emulate. In particular, Mt. Fuji-type landscapes cannot describe reciprocal sign epistasis, in which the presence of one mutation *a* changes whether another mutation *b* is beneficial, and vice versa, creating multiple optima (10).These non-additive 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 (4; 11), in which the system can be solely described by two parameters: the number of sites *N*, and *K*, the epistatic degree (the number of other sites influencing the effect of a given site). When *K*=0, the *NK* model gives a Mt. 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, reshuffling the genomic deck (12). 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 (13). In the block (or domains) model, mutations in one block only affect 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 only take 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 (14; 15).

Since its initial application to the maturation of the immune response (11), the *NK* model has been used to describe experimental protein and DNA fitness landscapes (16-

18). 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 to the experimental system (17; 18). 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*~0, attempts to fit the *NK* model to landscapes over wide regions might result in artificially low values of *K* due 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. 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 ref. (19).

Tunable ruggedness can also be introduced into the Mt. Fuji model (20; 21). The 'Rough Mt. Fuji' model is the addition of a Mt. 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 2a middle), 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.

3. Case study on evolutionary optimization: Neutral vs. frustrated networks

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 (22), 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 (23). Second, for common folds, the large set of sequences that share a given fold would form an evolutionary network throughout sequence space (24-27). 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 (~1/*N*), the islands become connected and the landscape as a whole exhibits a neutral network (28). 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 non-intuitively vast number of potential connections (29). 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 (30) (Figure 2e). 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 $p \rightarrow \infty$, 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 (31), 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 vs. frustrated network would involve mapping the topography of a complete fitness landscape. However, due 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 (32-34). Several examples of protein evolution to produce new or altered

function were also known (e.g., ref. (35)). These efforts were surprisingly successful, suggesting that different functions could be nearby in sequence space, i.e., 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 high-throughput sequencing. The first complete fitness landscape for short RNA sequences (N=21) revealed very few viable evolutionary paths between different functional families (36). 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 (37). 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 under-reported 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.

4. Measuring molecular fitness landscapes with high-throughput techniques

4A. RNA and DNA: from microarrays to HTS

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) (38), 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 (39), used with rational truncation to investigate the importance of structural constraints on aptamer activity (40), and combined with in silico approaches to interrogate large local evolutionary spaces in array-based directed evolution (41). A 2010 study 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 (17).

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 (42). In 2010, Pitt and Ferré-D'Amaré 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) (43). 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 (36). This requires a well-defined initial pool, but potentially expands the analysis, as it is no longer limited by the sequencing throughput 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 (44), and the prevalence of indirect evolutionary pathways that bypass local valleys (45) 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, un-measurable spaces (43; 46).

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 (36; 46), likely due to experimental noise and biases related to sequencing (e.g., PCR). Thus, new approaches use HTS to perform direct activity screens (47-49). Furthermore, fitness estimates can be notably improved by considering multiple rounds of selection (46).

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 (TF) (50). These data enabled mapping of DNA-TF binding energy over large sequence spaces (51), again illustrating the power of applying HTS to traditional questions.

4B. Beyond DNA and RNA: exploring new chemical space with HTS

Recent forays into the chemical space of nucleic acids (NAs) 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 NAs in a relatively short time (5). Alternative NAs raise many fundamental questions about fitness landscapes, from biologically inspired issues such as the uniqueness (or not) 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 NAs dates back to Eschenmoser's pioneering work (52), 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 (53; 54) or even two letters (55). In both cases, reduction in alphabet size led to selected ribozymes with lower activity than their larger-alphabet counterparts. On the other hand, artificially expanded genetic information systems (AEGIS) employ additional letters (56; 57) 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 an 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 (58-60).

For some functions, the activity of functional DNA molecules is comparable to that of RNA molecules (61). On some occasions, the sequence of a functional RNA can be simply synthesized as DNA and retain functionality (62; 63), sometimes requiring additional evolution (64). These exceptional cases may arise if the major interactions are electrostatic or nonspecific stacking interactions. XNAs made from non-natural

backbone alterations (Figure 4) have been selected for binding and catalytic activity, with activities similar to those seen in natural nucleic acids (65-67). Introduction of phosphorodithioate linkages can improve aptamer binding (68), with a single modified linkage increasing affinity by ~1,000-fold in one case (69). Another aspect of fitness is the chemical and physiological stability of the molecule; for example, many backbone modifications confer resistance to ribonuclease degradation (70). Other modifications, such as 2'-fluoro and 2'-amino RNA, provide both added stability (71) and sometimes increased functionality (72). The employment of chemical modifications to improve nucleic acids has been reviewed in more detail in ref. (73-75).

The application of HTS to alternative NAs is not trivial due to 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 (56; 65; 66). 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.

5. Fitness landscapes of organisms: RNA, proteins and genomes

Complete coverage of sequence space for an organismal genome – or even a single gene – is intractable due 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 (~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 (non-coding) RNAs (tRNA and snoRNAs) 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 (76; 77). As more RNA fitness landscapes are examined, it will be interesting to compare landscape characteristics of highly evolved biological RNAs vs. RNAs evolved in vitro to understand how >3.5 billion years of natural selection has shaped the landscape itself. Furthermore, the introduction of modified bases into cells (78) suggests the intriguing possibility of measuring fitness landscapes of alternative NAs 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 ~300, compared to ~150 for a typical ribozyme of length 50). 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 (79) (e.g., by cell sorting or simply by reproduction and survival for influenza variants (80)); DMS has been further reviewed (81; 82). 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 (83). 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 (84). On the other hand, DMS of a complete nine-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 (85). 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 (86).

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 (6). Weinreich et al. 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 (87), 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 (45). 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 Mt. Fuji model (88). 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.

6. 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 (89)) as well as sophisticated functions (e.g., ribozyme activity (90-92)). 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 E. coli, in which changes to the genetic background ('potentiating' mutations) enabled evolution of the ability to metabolize citrate,(93). 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 structure (94). Systematic study of the effect of the environment on the fitness landscape using HTS represents a major goal for this field.

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 (95). Crowded and confined conditions can modify the structure and function of nucleic acids and proteins (96-100). High levels of molecular crowding have been shown to stabilize mutations in ribozymes (101), change the binding mechanism of a ligand to a riboswitch (102), and create a chaperoning effect to assist in aptamer folding (99). Ribozymes can also modify their environment (e.g., through cooperation (103)), 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 (104; 105). 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 (106), and the accessible evolutionary pathways in an esterase were shown to change at different growth temperatures (107). Interestingly, variation in hosts may alter the topology of a viral fitness landscape, which may drive virus specialization (108). 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 (109). Further work in the yeast tRNA system also indicates that epistatic effects between loci can vary significantly for the same gene between different organisms (110). 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 (111). The "environmental landscape" for a single sequence can also be measured, as was done for a riboswitch in nearly 20,000 different environmental conditions (112). Measurement of such environmental

landscapes in conjunction with fitness landscapes is a challenging but essential goal for which high-throughput techniques are essential.

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 NCBI 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 still closer to reality.

Acknowledgements

Funding from the Simons Foundation (grant no. 290356), NASA (grant no. 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.

1. Luksza M, Lassig M. 2014. *Nature* 507:57-61
2. Wright S. 1932. *Proceedings of the Sixth International Congress of Genetics* 1:356--66
3. Smith JM. 1970. *Nature* 225:563-4
4. Kauffman S, Levin S. 1987. *Journal of theoretical biology* 128:11-45
5. Pinheiro VB, Holliger P. 2014. *Trends Biotechnol* 32:321-8
6. de Visser JA, Krug J. 2014. *Nat Rev Genet* 15:480-90
7. Obolski U, Ram Y, Hadany L. 2018. *Rep Prog Phys* 81:012602
8. Aita T, Husimi Y. 1996. *J Theor Biol* 182:469-85
9. Gillespie JH. 1983. *The American Naturalist* 121:691-708
10. Poelwijk FJ, Tanase-Nicola S, Kiviet DJ, Tans SJ. 2011. *J Theor Biol* 272:141-4
11. Kauffman SA, Weinberger ED. 1989. *J Theor Biol* 141:211-45
12. Kingman JFC. 1978. *Journal of Applied Probability* 15:1--12
13. Perelson AS, Macken CA. 1995. *Proc Natl Acad Sci U S A* 92:9657-61
14. Geard N, Wiles J, Hallinan J, Tonkes B, Skellett B. A comparison of neutral landscapes - NK, NKp and NKq. *Proc. Evolutionary Computation, 2002. CEC '02. Proceedings of the 2002 Congress on*, *2002*, 1:205-10:
15. Barnett L. 1998. In *Proceedings of the sixth international conference on Artificial life*:18-27. Madison, Wisconsin, USA: MIT Press. of 18-27 pp.
16. Hayashi Y, Aita T, Toyota H, Husimi Y, Urabe I, Yomo T. 2006. *PLoS One* 1:e96
17. Rowe W, Platt M, Wedge DC, Day PJ, Kell DB, Knowles J. 2010. *J R Soc Interface* 7:397-408
18. Franke J, Klozer A, de Visser JA, Krug J. 2011. *PLoS Comput Biol* 7:e1002134
19. Ferretti L, Schmiegelt B, Weinreich D, Yamauchi A, Kobayashi Y, et al. 2016. *J Theor Biol* 396:132-43
20. Neidhart J, Szendro IG, Krug J. 2014. *Genetics* 198:699-721
21. Aita T, Uchiyama H, Inaoka T, Nakajima M, Kokubo T, Husimi Y. 2000. *Biopolymers* 54:64-79
22. Kun Á, Szathmáry E. 2015. *Life (Basel, Switzerland)* 5:1497-517
23. Fontana W, Konings DA, Stadler PF, Schuster P. 1993. *Biopolymers* 33:1389-404
24. Tacker M, Fontana W, Stadler P, Schuster P. 1994. *European Biophysics Journal* 23:29-38
25. Huynen MA. 1996. *Journal of Molecular Evolution*
26. Fontana W, Schuster P. 1998. *Science (New York, N.Y.)* 280:1451-5
27. Schuster P, Fontana W, Stadler PF, Hofacker IL. 1994. *Proc Biol Sci* 255:279-84
28. Gavrilets S. 2004. *Fitness Landscapes and the Origin of Species (MPB-41)*. Princeton University Press
29. Gavrilets S. 2014. *J Hered* 105 Suppl 1:743-55
30. Stadler PF, Happel R. 1999. *J Math Biol* 38:435-78
31. Derrida B. 1981. *Phys Rev B* 24:2613-26
32. Schultes E, Bartel DP. 2000. *Science* 289
33. Held DM, Greathouse ST, Agrawal A, Burke DH. 2003. *J Mol Evol* 57:299-308
34. Curtis EA, Bartel DP. 2005. *Nature structural & molecular biology* 12:994-1000
35. Hayashi Y, Sakata H, Makino Y, Urabe I, Yomo T. 2003. *J Mol Evol* 56:162-8
36. Jimenez JI, Xulvi-Brunet R, Campbell GW, Turk-MacLeod R, Chen IA. 2013. *Proc Natl Acad Sci U S A* 110:14984-9
37. Petrie KL, Joyce GF. 2014. *J Mol Evol* 79:75-90

38. Sanchez-Luque FJ, Stich M, Manrubia S, Briones C, Berzal-Herranz A. 2014. *Sci Rep* 4:6242
39. Katilius E, Flores C, Woodbury NW. 2007. *Nucleic Acids Res* 35:7626-35
40. Fischer NO, Tok JBH, Tarasow TM. 2008. *PLoS ONE* 3:e2720-e
41. Knight CG, Platt M, Rowe W, Wedge DC, Khan F, et al. 2009. *Nucleic Acids Res* 37:e6
42. Athavale SS, Spicer B, Chen IA. 2014. *Current opinion in chemical biology* 22:35-9
43. Pitt JN, Ferre-D'Amare AR. 2010. *Science* 330:376-9
44. Otwinowski J, Plotkin JB. 2014. *Proceedings of the National Academy of Sciences* 111:E2301-E9
45. Wu NC, Dai L, Olson CA, Lloyd-Smith JO, Sun R. 2016. *eLife* 5
46. Pressman A, Moretti JE, Campbell GW, Muller UF, Chen IA. 2017. *Nucleic acids research* 45:10922
47. Kobori S, Yokobayashi Y. 2016. *Angewandte Chemie*
48. Dhamodharan V, Kobori S, Yokobayashi Y. 2017. *ACS Chem Biol* 12:2940-5
49. Jalali-Yazdi F, Lai LH, Takahashi TT, Roberts RW. 2016. *Angew Chem Int Ed Engl* 55:4007-10
50. Aguilar-Rodríguez J, Payne JL, Wagner A. 2017. *Nature ecology & evolution* 1:0045
51. Le DD, Shimko TC, Aditham AK, Keys AM, Longwell SA, et al. 2018. *Proceedings of the National Academy of Sciences*:201715888
52. Eschenmoser A. 1999. *Science* 284:2118-24
53. Rogers J, Joyce GF. 1999. *Nature* 402:323-5
54. Rogers J, Joyce GF. 2001. *RNA (New York, N.Y.)* 7:395-404
55. Reader JS, Joyce GF. 2002. *Nature* 420:841-4
56. Sefah K, Yang Z, Bradley KM, Hoshika S, Jiménez E, et al. 2014. *Proceedings of the National Academy of Sciences of the United States of America* 111:1449-54
57. Zhang L, Yang Z, Sefah K, Bradley KM, Hoshika S, et al. 2015. *J Am Chem Soc* 137:6734-7
58. Kimoto M, Yamashige R, Matsunaga K-i, Yokoyama S, Hirao I. 2013. *Nature Biotechnology* 31:453-7
59. Gawande BN, Rohloff JC, Carter JD, Von Carlowitz I, Zhang C, et al. 2017. *Proc Natl Acad Sci U S A* 114
60. Tolle F, Brändle GM, Matzner D, Mayer G. 2015. *Angewandte Chemie International Edition* 54:10971-4
61. Silverman SK. 2016. *Trends Biochem Sci* 41:595-609
62. Travascio P, Bennet AJ, Wang DY, Sen D. 1999. *Chem Biol* 6:779-87
63. Walsh R, DeRosa MC. 2009. *Biochem Biophys Res Commun* 388:732-5
64. Paul N, Springsteen G, Joyce GF. 2006. *Chem Biol* 13:329-38
65. Pinheiro VB, Taylor AI, Cozens C, Abramov M, Renders M, et al. 2012. *Science* 336:341-4
66. Taylor AI, Pinheiro VB, Smola MJ, Morgunov AS, Peak-Chew S, et al. 2015. *Nature* 69:208-15
67. Yu H, Zhang S, Chaput JC. 2012. *Nature chemistry* 4:183-7
68. Volk DE, Yang X, Fennewald SM, King DJ, Bassett SE, et al. 2002. *Bioorganic Chemistry* 30:396-419
69. Abeydeera ND, Egli M, Cox N, Mercier K, Conde JN, et al. 2016. *Nucleic acids research* 44:8052-64

70. Culbertson MC, Temburnikar KW, Sau SP, Liao JY, Bala S, Chaput JC. 2016. *Bioorg Med Chem Lett* 26:2418-21
71. Pieken WA, Olsen DB, Benseler F, Aurup H, Eckstein F. 1991. *Science (New York, N.Y.)* 253:314-7
72. Thirunavukarasu D, Chen T, Liu Z, Hongdilokkul N, Romesberg FE. 2017. *J Am Chem Soc* 139:2892-5
73. Dunn MR, Jimenez RM, Chaput JC. 2017. *Nature Reviews Chemistry* 1:0076
74. Rothlisberger P, Hollenstein M. 2018. *Adv Drug Deliv Rev*
75. Ni S, Yao H, Wang L, Lu J, Jiang F, et al. 2017. *Int J Mol Sci* 18
76. Li C, Qian W, Maclean CJ, Zhang J. 2016. *Science* 352:837-40
77. Puchta O, Cseke B, Czaja H, Tollervey D, Sanguinetti G, Kudla G. 2016. *Science* 352:840-4
78. Malyshev DA, Dhami K, Lavergne T, Chen T, Dai N, et al. 2014. *Nature* 509:385-8
79. Araya CL, Fowler DM. 2011. *Trends Biotechnol* 29:435-42
80. Phillips AM, Gonzalez LO, Nekongo EE, Ponomarenko AI, McHugh SM, et al. 2017. *Elife* 6
81. Fowler DM, Fields S. 2014. *Nat Methods* 11:801-7
82. Starita LM, Fields S. 2015. *Cold Spring Harb Protoc* 2015:711-4
83. Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, et al. 2010. *Nat Methods* 7:741-6
84. Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, et al. 2016. *Nature* 533:397-401
85. Hietpas RT, Jensen JD, Bolon DN. 2011. *Proc Natl Acad Sci U S A* 108:7896-901
86. Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, et al. 2012. *Nat Biotechnol* 30:543-8
87. Weinreich DM, Lan Y, Jaffe J, Heckendorn RB. 2018. *Journal of Statistical Physics*
88. Szendro IG, Schenk MF, Franke J, Krug J, De Visser JAG. 2013. *J Stat Mech-Theory E* 2013:P01005
89. Mimi G, Loana A, Roland W. 2017. *Angew. Chem., Int. Ed.* 56:2302-6
90. Attwater J, Wochner A, Holliger P. 2013. *Nat. Chem.* 5:1011
91. Schuabb C, Kumar N, Pataraia S, Marx D, Winter R. 2017. *Nat. Commun.* 8:14661
92. Frommer J, Appel B, Müller S. 2015. *Curr. Opin. Biotechnol.* 31:35-41
93. Blount ZD, Borland CZ, Lenski RE. 2008. *Proc Natl Acad Sci U S A* 105:7899-906
94. Melnikov A, Rogov P, Wang L, Gnirke A, Mikkelsen TS. 2014. *Nucleic Acids Res* 42:e112
95. Arroyo-Curras N, Dauphin-Ducharme P, Ortega G, Ploense KL, Kippin TE, Plaxco KW. 2018. *ACS Sens* 3:360-6
96. Saha R, Pohorille A, Chen IA. 2015. *Orig Life Evol Biosph* 44:319-24
97. Rivas G, Minton AP. 2016. *Trends Biochem. Sci.* 41:970-81
98. Mimi G, Christoph H, Satyajit P, Loana A, Gabriele S, Roland W. 2017. *ChemPhysChem* 18:2951-72
99. Saha R, Verbanic S, Chen IA. 2018. *Nature Communications* 9:2313
100. Daher M, Widom JR, Tay W, Walter NG. 2018. *J. Mol. Biol.* 430:509-23
101. Lee H-T, Kilburn D, Behrouzi R, Briber RM, Woodson SA. 2015. *Nucleic Acids Res.* 43:1170-6
102. Rode AB, Endoh T, Sugimoto N. 2018. *Angew Chem Int Ed Engl* 57:6868-72

103. Vaidya N, Manapat ML, Chen IA, Xulvi-Brunet R, Hayden EJ, Lehman N. 2012. *Nature* 491:72
104. Ho W-C, Zhang J. 2018. *Nat. Commun.* 9:350
105. Filteau M, Hamel V, Pouliot M-C, Gagnon-Arsenault I, Dubé AK, Landry CR. 2015. *Mol. Syst. Biol.* 11:832
106. Hietpas RT, Bank C, Jensen JD, Bolon DNA. 2013. *Evolution* 67:3512-22
107. Ota N, Kurahashi R, Sano S, Takano K. 2018. *Biochimie* 150:100-9
108. Cervera H, Lalic J, Elena SF. 2016. *J Virol*
109. Li C, Zhang J. 2018. *Nat Ecol Evol* 2:1025-32
110. Domingo J, Diss G, Lehner B. 2018. *Nature* 558:117-21
111. Qian W, Ma D, Xiao C, Wang Z, Zhang J. 2012. *Cell Rep.* 2:1399-410
112. Baird NJ, Inglese J, Ferré-D'Amaré AR. 2015. *Nature Communications* 6:8898-
113. Weinreich DM, Watson RA, Chao L. 2005. *Evolution* 59:1165-74
114. Pitt JN, Ferre-D'Amare AR. 2009. *Journal of the American Chemical Society* 131:3532-40
115. Humphrey W, Dalke A, Schulten K. 1996. *Journal of molecular graphics* 14:33-8, 27-8
116. Buenrostro JD, Araya CL, Chircus LM, Layton CJ, Chang HY, et al. 2014. *Nature Biotechnology* 32:562-8

Terms and Definitions list

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.

Alphabet - the set of chemical monomers ('letters') used in the construction of a biopolymer such as DNA or protein.

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

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

Fitness landscape - Function of fitness over sequence space.

Fitness Peak - Family of related sequences with elevated fitness.

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

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

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.

Neutral network - a network of evolutionary pathways on which fitness changes are negligible.

Frustration - Property describing a system that cannot simultaneously satisfy constraints of maximum favorability for each variable component.

In vitro selection (or evolution)- Laboratory evolution of biomolecules which selects sequences from a pool of variants based on ability to carry out a specific function.

Ribozyme - RNA sequence that catalyzes a specific reaction.

Aptamer - An RNA (or DNA) that binds to a specific ligand.

Riboswitch - Cis-acting RNA element containing an aptamer, in which ligand binding alters transcription or translation.

Reference Annotations

Ref. (113) - Perspective: Sign epistasis and genetic constraint on evolutionary trajectories.

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

Ref. (27) - From sequences to shapes and back: a case study in RNA secondary structures.

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

Ref. (43) - Rapid construction of empirical RNA fitness landscapes.

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

Ref. (17) - Analysis of a complete DNA-protein affinity landscape.

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

Ref. (36) - Comprehensive experimental fitness landscape and evolutionary network for small RNA.

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

Ref. (110) - Pairwise and higher-order genetic interactions during the evolution of a tRNA.

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

Ref. (6) - Empirical fitness landscapes and the predictability of evolution.

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

Ref. (81) - Deep mutational scanning: a new style of protein science.

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

Ref. (3) - Natural selection and the concept of a protein space.

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

Ref. (45) - Adaptation in protein fitness landscapes is facilitated by indirect paths.

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

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, (e) *N*=8. The fitness of each point in sequence space is represented by color (see legend) according to a smooth 'Mt. Fuji' landscape (e.g., fitness related to the number of '1's). 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 based on *Wright 1932* (2).

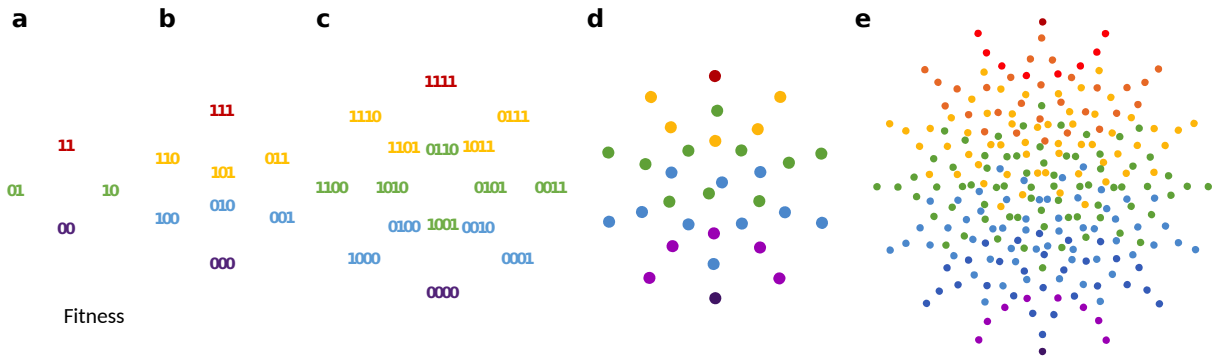
Figure 2. Epistasis and ruggedness on a fitness landscape. (a) For the simplest possible case (*m*=2, *N*=2), a smooth landscape can be climbed upwards from 00 to 11 (peak). Sign epistasis prevents passage over one trajectory, and reciprocal sign epistasis blocks both pathways (113). Fitness increase 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 1c). (c) A conceptual 3D depiction of fitness landscapes with varying ruggedness; horizontal axes correspond to sequence space and 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* (or *p*), conditions leading to maximum fitness (or minimum energy) conflict with each other and frustrate optimization.

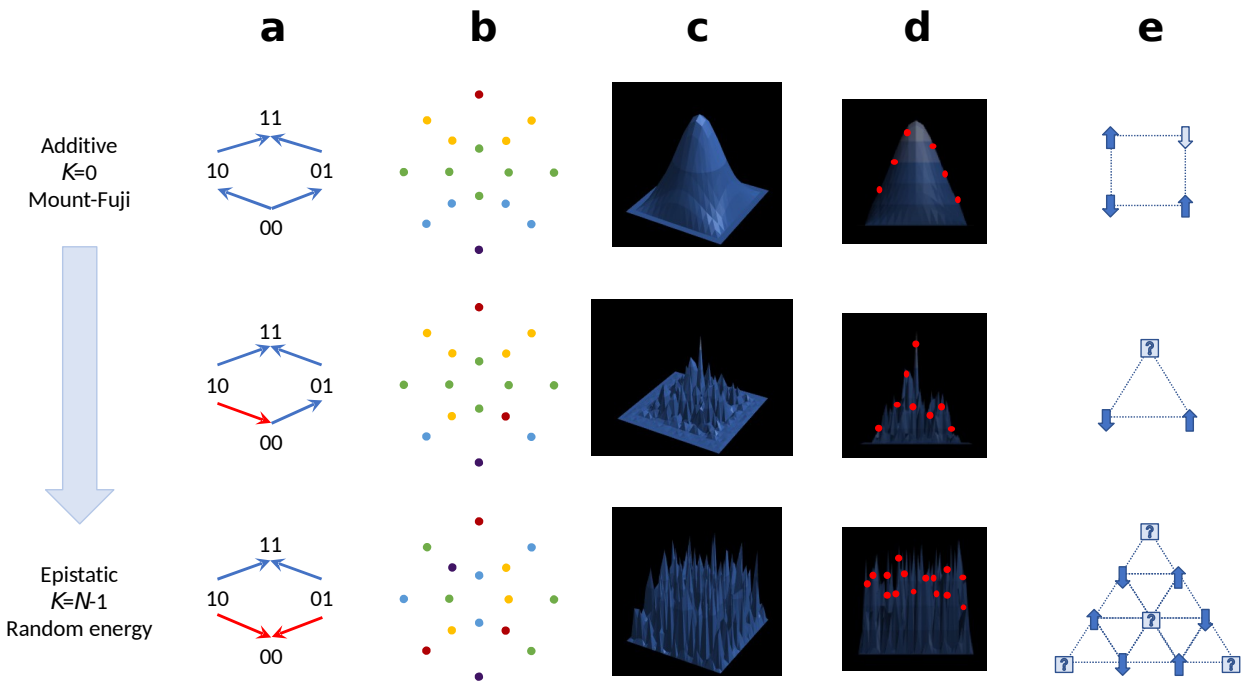
Figure 3. Stereo view of the structure of the class II ligase ribozyme (114) (PDB ID: 3FTM; image created with Visual Molecular Dynamics (115)).

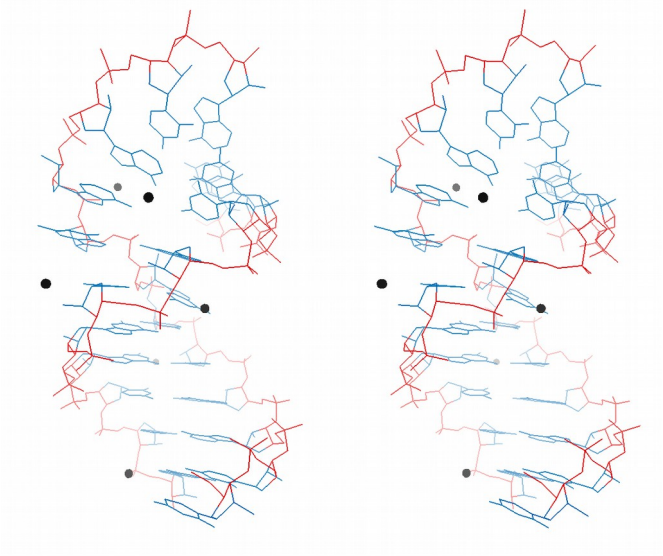
Figure 4. Expanded chemical space of functional nucleic acids. (a) The modified bases Ds (7-(2-thienyl)imidazo[4,5-b]pyridine) and EU (C5-ethynyl-uracil), which is utilized in click-SELEX. (b) Chemical structures for RNA (ribonucleic acid), DNA (deoxyribonucleic acid), 2'-F RNA (2'-fluoro RNA), ANA (arabino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

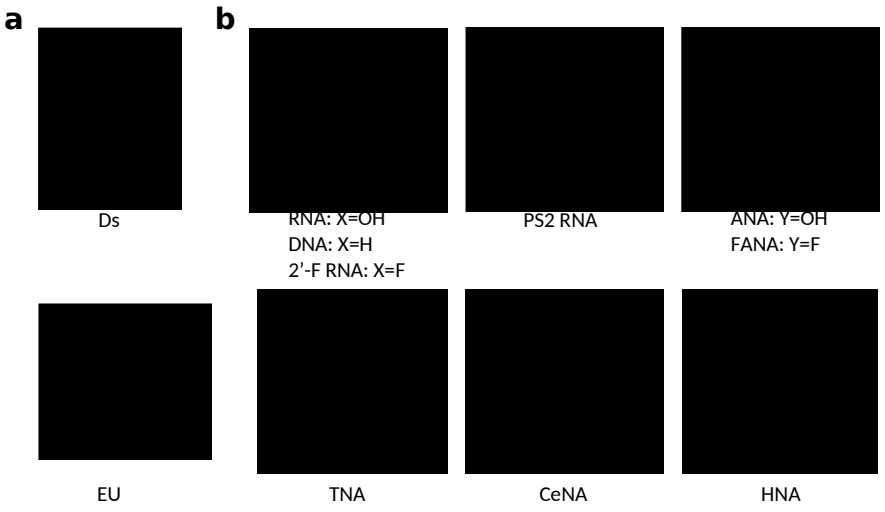
Figure 5. Representing HTS data of fitness landscapes. (a) A fitness peak with sequence space collapsed onto one dimension representing the edit distance (i.e., number of mutations) from the optimum sequence (after ref. (17; 36; 43)). (b) Evolutionary pathways between one local optimum and other nearby local optima, with sequence space collapsed as in (a). This representation illustrates fitness valleys and ruggedness (after ref. (17; 36)). (c) Heat map representing combinations of mutants, revealing epistatic interactions along the length of a sequence (after ref. (77; 116)).

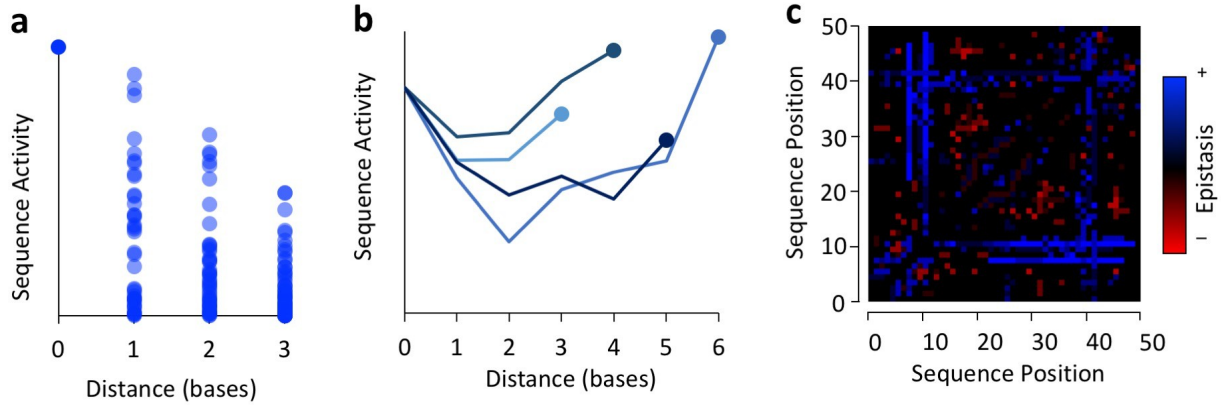
Figure 6. The fitness landscape depends strongly on the environment. For molecular fitness landscapes, environments might confer (a) stabilization of weakly folded structures (chaperoning), (b) exaggeration of fitness differences under stressed conditions, or (c) completely different structure in a new environment. The illustrations indicate the fitness landscape in one environment (dotted line) and in a new environment (solid line).

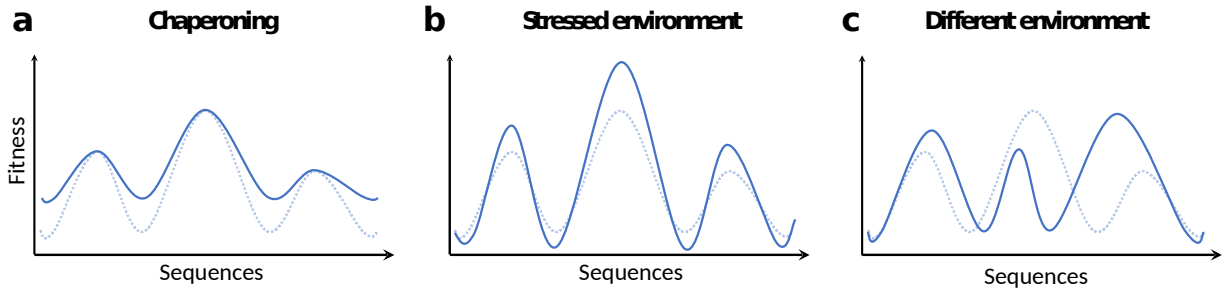



a
b
c
d
e
Additive
K=0
Mount-Fuji
11
10
01
00
Epistatic
K=N-1
Random energy




a
b
Ds
RNA: X=OH
DNA: X=H
2'-F RNA: X=F
PS2 RNA
ANA: Y=OH
FANA: Y=F
EU
TNA
CeNA
HNA


a
Sequence Activity
0
1
2
3
Distance (bases)
b
Sequence Activity
0
1
2
3
4
5
6
Distance (bases)
c
Sequence Position
50
40
30
20
10
0
0
10
20
30
40
50
Sequence Position
+
Epistasis
-


a
Chaperoning
Fitness
Sequences
b
Stressed environment
Sequences
c
Different environment
Sequences