Bridging the physical scales in evolutionary biology: from protein sequence space to fitness of organisms and populations

Shimon Bershtein[1], Adrian WR Serohijos[2] and
Eugene I Shakhnovich[3]

Bridging the gap between the molecular properties of proteins and organismal/population fitness is essential for understanding evolutionary processes. This task requires the integration of the several physical scales of biological organization, each defined by a distinct set of mechanisms and constraints, into a single unifying model. The *molecular* scale is dominated by the constraints imposed by the physico-chemical properties of proteins and their substrates, which give rise to trade-offs and epistatic (non-additive) effects of mutations. At the *systems* scale, biological networks modulate protein expression and can either buffer or enhance the fitness effects of mutations. The *population* scale is influenced by the mutational input, selection regimes, and stochastic changes affecting the size and structure of populations, which eventually determine the evolutionary fate of mutations. Here, we summarize the recent advances in theory, computer simulations, and experiments that advance our understanding of the links between various physical scales in biology.

Addresses
[1] Department of Life Sciences, Ben-Gurion University of the Negev, Beer-Sheva 84501, Israel
[2] Département de Biochimie, Centre Robert-Cedergren en Bioinformatique & Génomique, Université de Montréal, Montréal, QC H3T 1J4, Canada
[3] Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, United States

Corresponding author: Shakhnovich, Eugene I
(eugene@belok.harvard.edu)

Introduction

The direction of evolution is shaped by the interplay of mutations and selection. These two factors operate on vastly different scales: while mutations perturb the biophysical properties of proteins, potentially altering their structure, function and intracellular abundances, selection operates *directly* on the phenotypes and its outcome depends on fitness of organisms and populations. Understanding the link between the molecular effect of mutations and their fitness effects, the genotype-phenotype gap, is a major problem in biology. The relationship between the biophysical properties of macromolecules and their fitness effects is complex [1] and is influenced by various mechanisms operating at the *molecular*, *systems*, *organismal*, and, finally, *population* scales [2,3,4●●,5,6,7●,8●,9–14]. Although each scale of biological organization is usually studied independently by traditional scientific disciplines, it is becoming clear that more accurate understanding of evolutionary processes will depend on integrating these scales into a unified model [10].

A useful concept to visualize the linkage between genotype and phenotype in biology is *fitness landscape*. The concept was originally introduced by Sewall Wright in the early 1930s [15]. Wright's landscape can be easily imagined in three dimensions (in reality it is highly multidimensional) where axes x and y represent allele frequencies at loci 1 and 2, while axis z (the height) represents the mean fitness of a population. In such a representation, peaks and valleys describe, respectively, high and low mean fitness of a population. Selection is viewed as a force driving the allele frequencies to increase a population's mean fitness over time. However, if the landscape is rugged, that is, characterized by multiple peaks, selection is most likely to push a population to the nearest peak (local optimum), thus, effectively, locking it in a suboptimal fitness [16]. Climbing to the highest peak of the landscape (the global optimum) would require passing through a fitness valley that is disfavoured by selection.

The common use of the 'fitness landscape' metaphor often involves *discrete* values of fitness that correspond to specific alleles [16–18], or, more generally, discrete sequence variants [17,18]. However, such representations are limited to known variants, and their predictive power of the effects and evolutionary consequences of *de novo* mutations is limited. Conversely, full representation of the fitness landscape in sequence space is still very challenging due to high dimensionality of such mapping.

It is important to note that the term 'fitness' is often used in a variety of contexts that might be different from the traditional use of the term by Wright who defined fitness as a quantity that is proportional to the mean number of produced viable and fertile progeny [19]. For example, 'fitness' can stand in the literature for an organismal phenotype, systems level response, biological function of a protein, and, in case of sequence space-based landscapes, a molecular property that directly affects a protein's biological function [20–22]. Thus, the concept of 'fitness' is often interpreted to mean functional rather than reproductive capacity. However, in our view, such a broad definition of fitness detracts from the main conundrum in evolution that mutations and selection are separated by several levels of biological organization. To that end, we will exclusively use the term 'fitness' in the traditional sense close to Wright's as a phenotypic trait that is under selection and, as such, determines the outcome of a competitive evolutionary scenario in the wild, or in laboratory experiments. In the subsequent discussion we focus on organism-based definition of fitness as a reproductive capacity or related phenotypic traits of an organism linked to its specific genotype rather than a more traditional Wright's view of fitness as mean reproductive capacity of a population. We also note that specific phenotypic traits that are most relevant for the outcome of a competition might depend on the environment and ecological scenarios under which the competition occurs.

Another challenge that limits our ability to predict the direction and outcomes of evolution is that structure of a population has a crucial effect on how it evolves on the fitness landscape. Theoretical population genetics predicts that population size controls the balance between the forces of selection and random genetic drift; this eventually determines the direction of evolution on the fitness landscape. Earlier studies appreciated the role of population size, deriving an 'effective population size' to fit the genetics or the evolutionary data to existing population genetics models [23]. Nevertheless, direct experimental evidence for this fundamental concept is very scarce due to the inherent difficulty of simultaneously controlling the population size and obtaining tractable molecular readouts of evolutionary processes in a configurable and reproducible environment of a laboratory experiment, or in multiscale organism-based simulations.
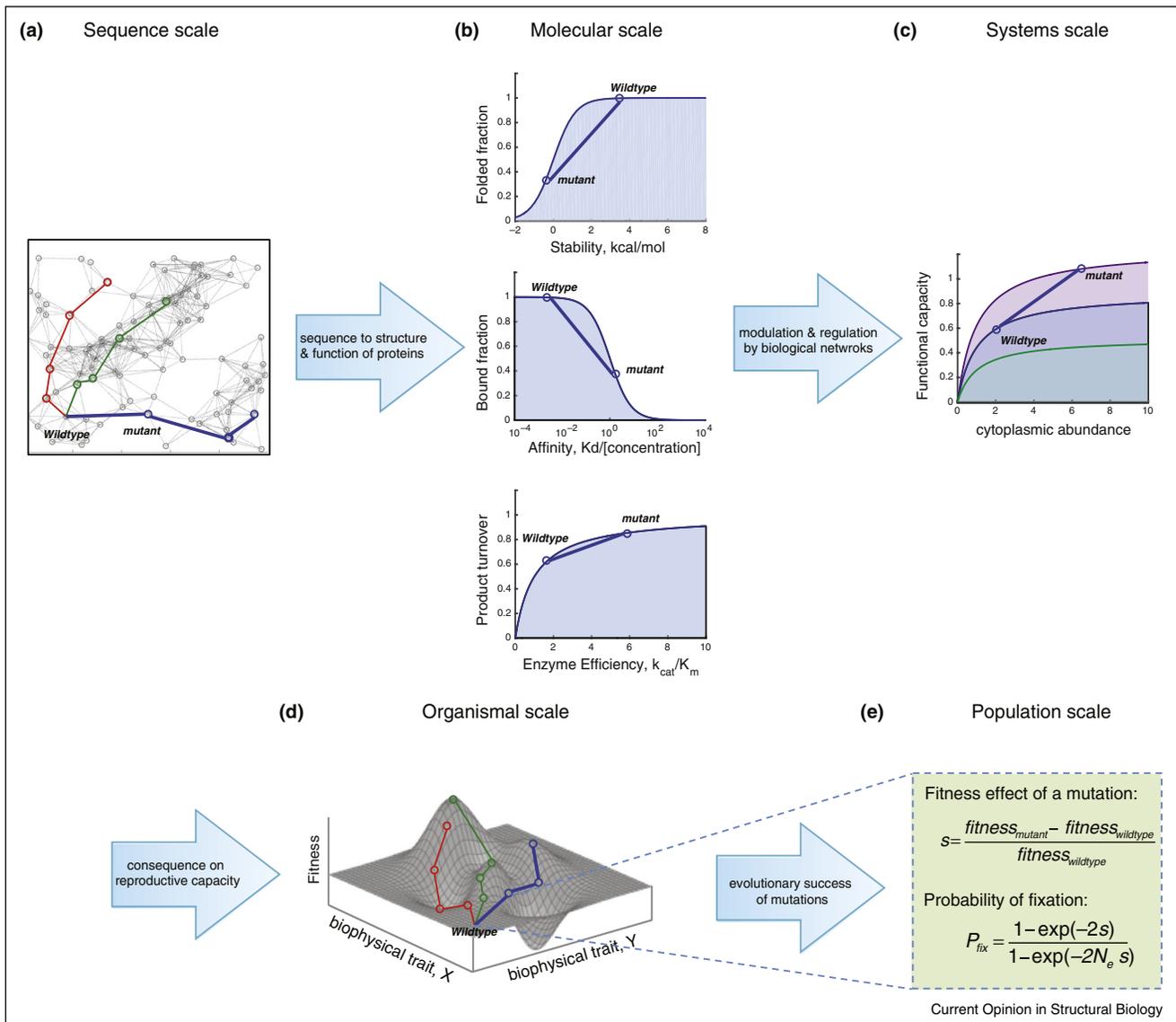
To address these challenges, several authors have recently introduced the concept of a 'biophysical fitness landscape'. The key premise of a biophysical fitness landscape is that the gap between genotype and fitness is bridged through the intermediate phenotype-molecular biophysical properties of proteins (Figure 1). In this approach the sequence-fitness gap might be overcome in two (or more) steps: (A) from sequence variation to variation of the molecular and systems-level properties of proteins (panels a–c in Figure 1) and (B) from variation of the molecular (stability, activity, etc.) and systems-level properties (e.g. abundances) to the organismal fitness effects reflected in a biophysical fitness landscape (panel d in Figure 1) to the fate of a mutation in a population (fixation or loss, panel e in Figure 1). The key assumption of this approach is that the complexity of mapping sequence variation to changes in the molecular properties (step A above, from panel a to panels b,c in Figure 1)) is at the root of the complexity of the sequence-fitness relationship ('ruggedness of fitness landscape'). Molecular effects of mutations are complex, degenerate and epistatic: similar molecular effects can be caused by different mutations and they strongly depend on the molecular background [9,22,24,25•]. However, at the next level(s) (step B), the relationship between changes in the molecular properties (stability, activity, aggregation propensity), cellular (intracellular abundance), and the ensuing phenotypic effects might be much less degenerate and predictive, giving rise to a smooth fitness landscape in the space of molecular properties of proteins (Figure 1). It is also important to note that pleiotropy, when mutations have conflicting effects on different molecular traits, is another major source of epistasis. In terms of the biophysical fitness landscape this means that projection of individual mutational effects could lead to complex trajectories in the space of biophysical parameters. In this review we present recent theoretical, computational and experimental studies that elaborate on and further develop these ideas.

From protein sequence space to protein biophysics

Maynard Smith was first to suggest that evolution can be viewed as a walk through a protein sequence space [26]. Assuming an L-residue long protein sequence, its sequence space will constitute a network of 20^L sequences interconnected through edges, each representing a change in a single residue. Each of the sequences is assigned a value (a molecular property such as thermodynamic stability, or function), thus forming a discrete landscape of the sequence space. Because mutations are rare, evolutionary trajectories are thought to traverse the sequence space by single mutation steps, without passing through non-functional intermediates. The resulting accessible fraction of the interconnected *functional* sequences constitutes the (nearly) neutral protein network [27–30]. If the effect of each mutation on fitness is independent on the genetic background on which it occurs, the resulting fitness landscape will contain a single peak populated by optimal sequences with multiple evolutionary trajectories leading to it. Conversely, should epistasis (non-additivity of mutations) be prevalent, fitness effect of a mutation might be beneficial or detrimental depending on the genetic background [31,32•]. Epistasis can severely constrain the accessible evolutionary pathways, and create a rugged fitness landscape with

Figure 1

(a) Sequence scale **(b)** Molecular scale **(c)** Systems scale

sequence to structure & function of proteins

modulation & regulation by biological netwroks

(d) Organismal scale **(e)** Population scale

consequence on reproductive capacity

evolutionary success of mutations

Fitness effect of a mutation:

$$s = \frac{fitness_{mutant} - fitness_{wildtype}}{fitness_{wildtype}}$$

Probability of fixation:

$$P_{fix} = \frac{1 - \exp(-2s)}{1 - \exp(-2N_e\,s)}$$

Current Opinion in Structural Biology

Biophysics as a stepping stone between sequence and phenotype. Closing the genotype-phenotype gap is facilitated by an intermediate projection of organismal fitness to biophysical properties of macromolecules. While the effect of sequence variation (panel **a**) on molecular traits (e.g. folding stability, binding affinity, catalytic activity, panel **b**) might be complex, the ensuing relation between variation of the observable biophysical traits (vertical axes in panel b) and their fitness effect on the organism might be simple and predictable in some cases (panel **d** and Figure 2). The effects of mutations are also modulated by a regulation of biological networks (panel **c**) that might also have simple integrative effect on fitness. On the level of populations, the probability of fixing a specific mutation is a function of the effect of a mutation (selection coefficient) and an effective population size (N_e). Note a different interpretation of fitness landscape in (panel d) from classical Wright's where mean fitness of the population versus allele frequencies is usually plotted.

multiple local optima separated from each other and the global fitness peak by trajectories passing through non-functional sequence states [33[••],34,35]. Kaltenbach *et al.* [33[••]] examined the evolutionary reversibility of phosphotriesterase that was evolved by directed evolution to arylesterase and back, and found that many of the original amino acid exchanges between the new and the original function could not be tolerated, and an alternative set of mutations was needed to restore the phosphodies-terase activity. Extensive epistasis made the adaptive fitness landscape *in sequence space* highly rugged, and, although the evolutionary trajectories were phenotypical-ly (at the level of protein phenotype, function) reversible, the genotypic trajectory became irreversible, suggesting that the new and the original activities constitute separate fitness peaks [36].

Which biophysical properties determine the epistatic interaction between mutations? Various mechanisms were proposed, including protein-ligand or protein-DNA binding [25•,37,38••], protein conformation [33••,39], and allostery [40]. However, protein stability appears to be the most prevalent mechanism of intramolecular epistasis [7•,41••,42,43]. Since most mutations destabilize protein structures [14,44], it was asserted that protein evolution is often accompanied by stability-activity trade-offs [45]. Indeed, Gong *et al.* [41••] demonstrated that evolution of influenza nucleoprotein is constrained by a stability-related epistasis: acquisition of stabilizing mutations was required before obtaining the adaptive substitutions, which, on their own, caused adverse effects due to destabilization of the nucleoprotein and, therefore, were evolutionary inaccessible. Stability activity trade-off is often cast as the effect of stabilizing mutations on 'flexibility' of the protein which is detrimental to 'functional protein dynamics' [46,47]. Experimental studies indeed showed that mutations introduced *in the active site of a protein* stabilize the protein yet make it less or completely inactive [48]. However, more comprehensive analyses where mutations were introduced throughout the protein, and not just in the active site, showed no inverse relation between stability and activity [36,49]. Moreover, a weak *positive* correlation between activity and stability was observed for multiple variants of dihydrofolate reductase (DHFR) from *Escherichia coli* [36]. Nonetheless, a highly stabilizing mutation in the active site of DHFR, D27F, completely eliminated its function [36], similar to what was observed for another enzyme in [48]. The apparent disparity of the conclusions from substitutions in/near the active site and elsewhere in the protein can be rationalized by the observation that most 'random' mutations in proteins are destabilizing [13,44,50]. Carving an active site on an enzyme requires several functional substitutions, which, from the point of view of an independent trait – stability – could be random or even detrimental (e.g., placing charged residues in partly buried areas of the protein). Thus, activity-stability trade-off could be real for substitutions in the active sites. However, it disappears or reverses itself for substitutions outside the active sites. Two possibilities can be raised to explain this finding: (1) that global stability of a protein against unfolding might not be relevant for its global dynamics in the native state and (2) that global dynamics might not determine functionality, as was indeed argued in [51]. To illustrate these arguments, the following analogy with building a house might be offered. In order for a house to be livable, it must have windows. Windows certainly diminish structural integrity of the building (decreased 'stability'). However, that does not mean that a building must be structurally shaky to be livable. What it does imply, though, is that other parts of the building must be reinforced to allow for windows to be carved in the walls without the building falling apart completely — the analogy with the need to make stabilizing mutations elsewhere in the protein

to evolve functional variants [49,52]. A strong evidence that overall stabilization of proteins does not come at a detriment to its function comes from long term evolutionary experiments with *E. coli* [53]. Most lines that evolved at 20 °C lost fitness relative to the ancestor when they competed at 40 °C and above [88], whereas most lines that evolved at 41.5 °C did not lose fitness at 20 °C and below [54].

The prevalence of epistasis, and its impact on evolution is still debated. Some describe it as rampant [38••] and pervasive [55], or even consider it a determining factor of the molecular evolution [22,56], whereas others view it as less important [57,58]. We return to the discussion of prevalence and mechanisms of epistasis later, after introducing the theoretical foundation of protein biophysics-population genetics mapping.

From protein biophysics to organismal fitness

Although insightful and informative, evolutionary studies that focus entirely on the protein sequence space landscapes have one major disadvantage — being purely phenomenological, they are oblivious to the biophysical mechanisms of the molecular and systems scales that determine fitness effects of mutations [59]. Indeed, biological networks are responsible for the striking capacity of organisms, on one hand, to limit the enormous fitness cost of the mutational load [60,61], and, on the other hand, to harness its evolutionary potential [62]. A major role in modulating the mutational effects at a systems scale is played by protein quality control (PQC). In general, molecular chaperones constituting the PQC can serve as a buffer against the deleterious effects of mutations, thus they are less likely to be purged by purifying selection [12,63]. This lead to the view that certain elements of the PQC act as 'capacitors' of genetic variability that broadly reshape the biophysical properties — organismal fitness map [12,64–67] and control the rate of protein evolution [63,68–72]. Recently, PQC was also identified as imposing a global barrier to functional integration of horizontally transferred genes in bacteria, because the newly acquired genes were found to be incompatible with the elements of PQC of the host [73••].

Rodrigues *et al.* [4••] have recently demonstrated that it is possible to incorporate the pleiotropic effects of mutations originating in the PQC interaction with the mutant proteins into a model that correctly predicts organismal fitness entirely from the biophysical characteristics. The model is based on the relationship between fitness and metabolic flux [74,75]:

$$\text{fitness} \sim \text{flux} = \frac{a \cdot E}{B + E}, \tag{1}$$

where E is the functional capacity of a protein in the enzymatic chain, a denotes maximal fitness at the highest flux, and B is the constant related to the effect of all other

proteins in an enzymatic chain. The functional capacity is proportional to the product of the net intracellular abundance of the functional protein molecules (E_0) and its catalytic efficiency (k_{cat}/K_M):

$$E \sim E_0 \frac{k_{cat}}{K_M}. \qquad (2)$$

In the presence of a competitive inhibitor Eq. (2) can be further modified as:

$$E_{eff} \sim E_0 \frac{k_{cat}}{K_m} K_i, \qquad (3)$$

where K_i is the enzyme inhibition constant and E_{eff} is the effective functional capacity.

Rodrigues *et al.* [4••] chose DHFR as a convenient model, since bacterial fitness can be accurately predicted from the metabolic flux through DHFR according to Eq. (1) [2,12,73••] (Figure 2). They showed that mutations conferring trimethoprim resistance in DHFR are highly pleiotropic, that is, simultaneously affecting multiple molecular traits: catalytic efficiency, trimethoprim binding K_i, and protein stability, thus rendering the biophysical fitness landscape multidimensional. While k_{cat}/K_m and K_i can be measured *in vitro*, the successful implementation of the model also requires the knowledge of the functional intracellular abundance E_0 (Eq. (3)). Apart from the thermodynamic stability that determines the fraction of the folded protein molecules at equilibrium [76], protein abundance, E_0, also depends on the interaction with

PQC [12]. Rodrigues *et al.* [4••] showed that this latter quantity can be accurately predicted from the *in vitro* measurements of the population of the molten-globule-like state of the mutant proteins, in agreement with theoretical predictions from the dynamic turnover model [12]. Using the model, Rodrigues *et al.* [4••] successfully predicted the fitness effects (IC_{50}) of the trimethoprim-resistance conferring mutations entirely from the molecular properties of mutant DHFRs in a broad range of conditions.

The map between the biophysical to systems to organismal scales can also be assessed quantitatively by global transcriptomics and proteomics techniques [2,77,78]. Bershtein *et al.* [2] demonstrated that local perturbation of a central metabolic enzyme by chromosomal incorporation of destabilizing mutations produced a global perturbation of protein abundances for a large number of genes well beyond the local metabolic neighbourhood of the enzyme. An empirical quantitative relation was established between the biophysical properties of the mutants (change in thermodynamic stability), global proteome response to the destabilizing mutations (measured as standard deviations of the distributions of logarithms of relative to wild-type protein abundances), and bacterial fitness of the mutant strains (approximated as growth rates).

Altogether, these works, as well as the recent studies of viral fitness [41••], demonstrate the utility of biophysical fitness landscapes in providing the predictive quantitative multiscale relationships between variation of the

Figure 2
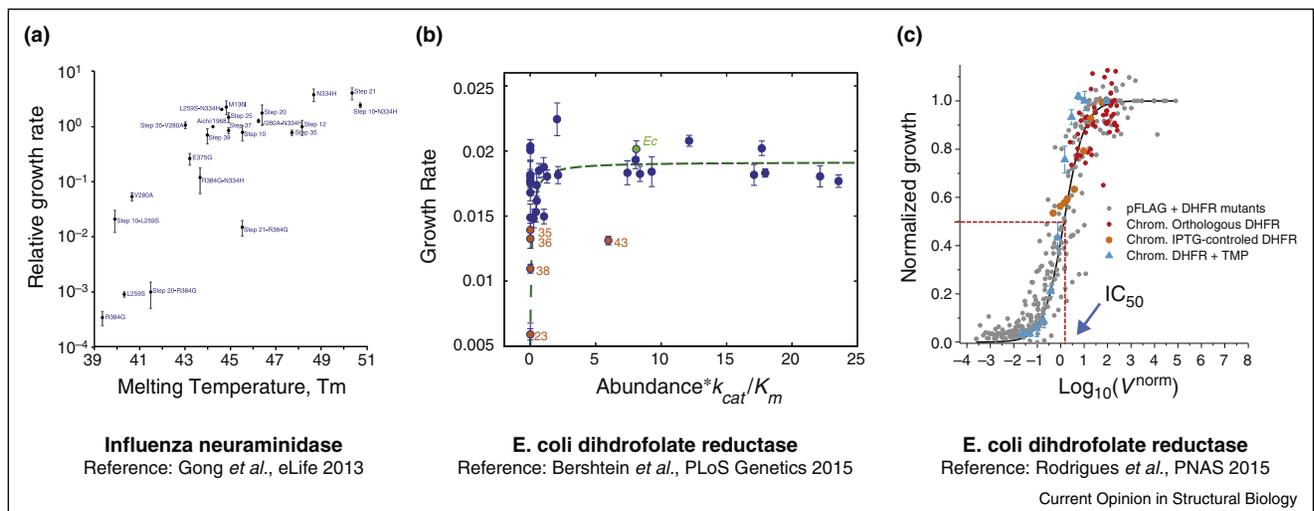
Experimentally derived biophysical fitness landscape in viruses and microorganisms. **(a)** Fitness of influenza from strain Aichi/1968 to strain Brisbane/2007 is mapped to the thermal stability of its neuraminidase domain [41••]. Viral fitness is defined as the number of particles that are able to productively infect cells and transcribe high levels of GFP from viral RNA. **(b)** Fitness of *E. coli* (defined as growth rate) is mapped to the effective functional capacity (Abundance*k_{cat}/K_m) of dihydrofolate reductase, a core metabolic enzyme. Points correspond to strains where the chromosomal copy of DHFR was replaced with an ortholog to mimic xenologous horizontal gene transfer [73••]. **(c)** Fitness of *E. coli* under antibiotic resistance (defined as IC50) is mapped to the rate of DHFR catalyzed reaction. Points correspond to strains/conditions where the *E. coli* DHFR has been mutated and/or its abundance has been titrated [4••].

molecular properties and the ensuing fitness effects. Figure 2 summarizes the recent examples of biophysical fitness landscapes where simple deterministic relationships between certain combinations of physical traits and fitness were established.

From protein biophysics to population genetics

Mapping of the observed molecular biophysical properties of proteins to evolutionary history of populations is one of the most challenging aspects of the evolutionary studies. The difficulty of theoretical analyses stems from the fact that factors affecting the fixation probabilities of mutations at a population scale, such as mutation rate, effective population size, drift, selection regimes, etc., are (*i*) stochastic and (*ii*) are of non-biophysical nature. The experimental difficulty is rooted, on one hand, in the inherent complexity of the living systems that makes it very difficult to disentangle the pure biophysical properties of the evolving proteins from the pleiotropic influences, and on the other hand, in the scarcity of the data on the evolutionary intermediates. Considerable progress in the field in the last decade [6,7•,9,14,50,79,80] came from the theoretical synthesis of the population genetics theory [81] with statistical thermodynamics of protein stability and folding [82].

Regardless of a specific function, globular proteins must be stable enough in order to preserve their native structure and function. Thus, from a molecular biophysics perspective, protein thermodynamic stability (described as Gibbs free energy difference between folded and unfolded states, ΔG), is one of the major determinants of sequence evolution [50,83–85]. Under a selection pressure to maintain protein stability, fitness of an organism, F, is assumed to be proportional to the number of the folded proteins: $F \sim P_{\mathrm{nat}}$, where P_{nat} is the probability to find protein in its native state at equilibrium, given the two-state model of protein folding [86]:

$$P_{\mathrm{nat}} = \frac{1}{1 + e^{\Delta G/kT}}. \tag{4}$$

The effect of an accrued mutation on protein stability is calculated as:

$$\Delta G_{\mathrm{after}} = \Delta G_{\mathrm{before}} + \Delta\Delta G_{\mathrm{mutation}}, \tag{5}$$

were $\Delta\Delta G_{\mathrm{mutation}}$ stands for the change in ΔG upon mutation, and can be directly estimated using force field based stability predictors [87,88], calculated by summing the contact energies of all of the pairs of contact-making residues in a conformation [89], or, in a more phenomenological approach, drawn from a Gaussian distribution with mean 1 kcal/mol, and SD = 1.7 kcal/mol [9,50], — parameters that are derived from empirical studies [90] and computational estimates [45]. The strong non-linearity of the Fermi-Dirac function Eq. (4) makes the fitness

effects of stability-changing mutations *inherently* epistatic: fitness effects of mutations are more pronounced on low stability backgrounds [9,76]. However, because of the assumed additivity of $\Delta\Delta G$ values [91], the model cannot account explicitly for the instances of site-site epistasis. To this end, additional terms correcting for the site-site epistasis in the energy function were suggested [11,43].

An arising mutation would have a selection coefficient, s, defined as [9,11]:

$$s = \frac{F_{\mathrm{after}} - F_{before}}{F_{\mathrm{before}}} \sim e^{\Delta G_{\mathrm{before}}/kT} e\big(1 + e^{\Delta\Delta G_{\mathrm{mutation}}/kT}\big). \tag{6}$$

Importantly, Eq. (6) allows establishing a direct link between a biophysical property of a protein afflicted by a mutation and the probability of fixation of the mutation (P_{fix}) in a *monoclonal* population with an effective population size N_{eff} [92]:

$$P_{\mathrm{fix}} = \frac{1 - e^{-2s}}{1 - e^{-2sN_{\mathrm{eff}}}}. \tag{7}$$

Eqs. (4)–(6) lay the theoretical foundation for theory and evolutionary simulations in the class of models where fitness is directly linked to stabilities of proteins encoded in genomes of organisms, and populations of such organisms are subjected to mutations, drift, and selection [6]. This approach provided important evolutionary insights into the distribution of fitness effects of mutations [9,80], the molecular determinants of the rate of protein evolution [93,94], the genetic variation in coding regions [79], and the observed thermodynamic and structural properties of proteins in nature [11,95]. The key insight from these studies is the realization that 'marginal' stabilities of natural proteins is a direct consequence of mutation-selection balance [11,95], rather than selection against an excessive protein stability. The theory based on mutation-selection balance quantitatively predicts the distribution of protein stabilities in a remarkable agreement with experiments [9,50]. Additionally, the integration of protein folding stability and population genetics leads to some major insights in molecular evolution. First, the selection for protein folding stability (or its contrapositive, selection against protein misfolding) could explain the genomics observations of the rate of protein evolution (dN and dS). These trends, observed across all kingdoms of life, include the observation that highly expressed proteins tend to evolve slowly [94,96,97], and the apparent universality of the log-normal distribution of evolutionary rates [98]. Second, selection for folding stability has also been shown to predominantly shape the pattern of polymorphisms in the protein coding regions [79]. Third, it has also been theoretically shown that on the biophysical fitness landscape defined by folding stability there exists a mathematical relationship between the folding stability of a protein, its abundance in the cell,

and the effective population size of the organism [9,95]. This mathematical relationship is important because, for the first time, it quantifies (in units of energy) the contribution of variables, such as population size and protein abundance, to the observed folding stability of proteins in nature. Altogether, these mechanistic insights on the rate of protein evolution and patterns of polymorphisms arising from integrating biophysics and population genetics could have fundamental practical application on the statistical tools used to infer selection and adaptive evolution of sequencing data. We note that inference of evolutionary forces based on observed genetic variation are largely based on the Neutral Theory (neutrality tests such as dN/dS, McDonald–Kreitman, Hudson–Kreitman–Aguade (HKA), etc.) that altogether ignores these mechanistic aspects of protein evolution.

Lastly, Eqs. (4)–(6) also provide for the simple and universal mechanism of epistasis on the biophysical fitness landscape. The mapping between the effect of mutations on folding stability ($\Delta\Delta G_{\text{mutation}}$) and the selection coefficient (s) depends on the wildtype (or background) folding stability ΔG_{before}. In particular, when the proteins are sufficiently stable ($\Delta G_{\text{before}} \ll -5$ kcal/mol), the factor $e^{\Delta G_{\text{before}}/k\text{T}}$ approaches zero, which then renders most mutations neutral ($s \sim 0$). Alternatively, when the proteins are close to being unfolded, the non-zero factor $e^{\Delta G_{\text{before}}/k\text{T}}$ modulates the effect of $\Delta\Delta G_{\text{mutation}}$ on s. Indeed, by running explicit evolutionary simulations on biophysical fitness landscape determined by folding stability, Eqs. (4)–(6) Serohijos et al. showed that epistasis is prevalent but generally weak [79]. Interestingly, by using deep mutational screening analysis on GB1 protein, a technique allowing a comprehensive characterization of pairwise epistasis [99,100], Olson et al. have arrived to a similar conclusion [42]. They showed that strong epistasis is relatively rare ($\sim$5% of pairwise interaction), whereas weak epistasis is widespread ($\sim$30%). Additionally, according to Olson et al. there is a strong epistasis among destabilizing mutations, but not among stabilizing mutations. This again finds an explanation in the biophysical fitness landscape based on protein folding thermodynamics. Two stabilizing mutations bring the protein to the flatter part of the landscape, hence less epistasis, but two destabilizing mutations shift it to the more curved part of the landscape, hence stronger epistasis. In a similar vein, Bloom and coauthors showed that stability-mediated weak epistasis of the type described in [79] is prevalent in evolution of influenza nucleoprotein [41••].

Recently, these ideas were extended to a multidimensional biophysical fitness landscape that encompasses several molecular traits. Interaction with other proteins is almost as universal a trait as protein folding. Cheron et al. studied a biophysical fitness landscape of viral proteins escaping antibody stressors [101]. They used a simple biophysics-based model of viral fitness which assumed that successful escape from antibody required that viral proteins maintain their folded state while decreasing the affinity of the antibody to the viral antigen and showed how viral demographics affects evolution of stability of the viral antigen and its binding to antibody. Further, they showed how evolutionary dynamics on biophysical fitness landscape determines optimal (for the viral escape) mutations rates. Another interesting example of the interplay between folding and binding on biophysical fitness landscape is given in [8•,102] where binding affinity to other proteins can evolve as an evolutionary spandrel to maintain stable folding of proteins.

Conclusions and outlook
The concept of biophysical fitness landscape is helping to advance our understanding of the interplay of mutation and selection in determining the course and outcomes of evolution [10]. Several examples of simple quantitative and predictive relationships between biophysical properties of proteins and fitness contributed to mechanistic understanding of how Darwinian selection at the population level shapes the evolution of molecular properties of enzymes [10]. The availability of quantitative biophysical fitness landscape opens up an exciting opportunity to develop accurate multiscale models to predict evolutionary dynamics from first biophysical principles. Important applications include bacterial and viral escape from stressors such as antibiotic, antibody [101] and thermal adaptation [103]. The challenges ahead include extension of the biophysical fitness landscape from metabolic enzymes to other functional classes, including accurate fitness models of gene expression and integrating systems-level global effects on metabolic and proteomic states of cells. We expect that combination of predictive biophysical models with advances in genomics and proteomics will bring quantitative *ab initio* prediction of evolution within the reach of many applications of fundamental and translational significance.

Conflict of interest statement
Nothing declared.

Acknowledgements

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Bershtein S, Mu W, Shakhnovich EI: **Soluble oligomerization provides a beneficial fitness effect on destabilizing mutations**. *Proc Natl Acad Sci U S A* 2012, **109**:4857-4862.

2. Bershtein S, Choi JM, Bhattacharyya S, Budnik B, Shakhnovich E: **Systems-level response to point mutations in a core metabolic enzyme modulates genotype–phenotype relationship**. *Cell Rep* 2015, **11**:645-656.

3. Lind PA, Tobin C, Berg OG, Kurland CG, Andersson DI: **Compensatory gene amplification restores fitness after inter-species gene replacements**. *Mol Microbiol* 2010, **75**:1078-1089.

4. Rodrigues JV, Bershtein S, Li A, Lozovsky ER, Hartl DL,
•• Shakhnovich EI: **Biophysical principles predict fitness landscapes of drug resistance**. *Proc Natl Acad Sci U S A* 2016, **113**:E1470-E1478.
This paper demonstrates that a biophysics-based fitness landscape of an essential central metabolism enzyme (DHFR) can provide an accurate prediction of the antibiotic resistant phenotypes. Specifically, a kinetic flux model was used to predict, with high accuracy, the IC50 of trimetho-prim resistance from a unique combination of the molecular (activity, binding, and folding stability) and cellular (intracellular abundance) para-meters of the mutant and orthologous forms of DHFR. Further, this work established the role of PQC in the evolution of antibiotic resistance.

5. Dasmeh P, Serohijos AW, Kepp KP, Shakhnovich EI: **Positively selected sites in cetacean myoglobins contribute to protein stability**. *PLoS Comput Biol* 2013, **9**:e1002929.

6. Serohijos AW, Shakhnovich EI: **Merging molecular mechanism and evolution: theory and computation at the interface of biophysics and evolutionary population genetics**. *Curr Opin Struct Biol* 2014, **26**:84-91.

7. Shah P, McCandlish DM, Plotkin JB: **Contingency and
• entrenchment in protein evolution under purifying selection**. *Proc Natl Acad Sci U S A* 2015, **112**:E3226-E3235.
This work uses computational models of thermodynamic stability and simulations of protein sequence evolution by sequential fixation of nearly neutral mutations to study epistasis, contingency, and entrenchment under long-term purifying selection on protein stability. Importantly, the paper demonstrates that fixed mutations often become entrenched by epistasis — substitutions that were nearly neutral upon fixation become increasingly deleterious to revert as subsequent substitutions accumulate.

8. Manhart M, Morozov AV: **Protein folding and binding can
• emerge as evolutionary spandrels through structural coupling**. *Proc Natl Acad Sci U S A* 2015, **112**:1797-1802.
By applying a biophysical and evolutionary model, the paper examines how the inherent structural coupling between protein stability and binding gives rise to an evolutionary coupling between them. The authors con-clude that these traits can co-evolve even though their coevolution does not confer a fitness advantage *per se* (so-called, evolutionary 'span-drels').

9. Wylie CS, Shakhnovich EI: **A biophysical protein folding model accounts for most mutational fitness effects in viruses**. *Proc Natl Acad Sci U S A* 2011, **108**:9916-9921.

10. Liberles DA, Teichmann SA, Bahar I, Bastolla U, Bloom J, Bornberg-Bauer E, Colwell LJ, de Koning AP, Dokholyan NV, Echave J, Elofsson A *et al.*: **The interface of protein structure, protein biophysics, and molecular evolution**. *Protein Sci* 2012, **21**:769-785.

11. Goldstein RA: **The evolution and evolutionary consequences of marginal thermostability in proteins**. *Proteins* 2011, **79**:1396-1407.

12. Bershtein S, Mu W, Serohijos AW, Zhou J, Shakhnovich EI: **Protein quality control acts on folding intermediates to shape the effects of mutations on organismal fitness**. *Mol Cell* 2013, **49**:133-144.

13. Zeldovich KB, Chen PQ, Shakhnovich EI: **Protein stability imposes limits on organism complexity and speed of molecular evolution**. *Proc Natl Acad Sci U S A* 2007, **104**:16152-16157.

14. Bloom JD, Raval A, Wilke CO: **Thermodynamics of neutral protein evolution**. *Genetics* 2007, **175**:255-266.

15. Wright S: **The role of mutation, inbreeding, crossbreeding, and selection in evolution**. *Proc 6th Int Cong Genet* 1932, **1**:356-366.

16. de Visser JA, Krug J: **Empirical fitness landscapes and the predictability of evolution**. *Nat Rev Genet* 2014, **15**:480-490.

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

18. Neher RA, Russell CA, Shraiman BI: **Predicting evolution from the shape of genealogical trees**. *eLife* 2014:3.

19. Orr HA: **Fitness and its role in evolutionary genetics**. *Nat Rev Genet* 2009, **10**:531-539.

20. Harms MJ, Thornton JW: **Evolutionary biochemistry: revealing the historical and physical causes of protein properties**. *Nat Rev Genet* 2013, **14**:559-571.

21. Soskine M, Tawfik DS: **Mutational effects and the evolution of new protein functions**. *Nat Rev Genet* 2010, **11**:572-582.

22. Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, Sharonov GV, Ivankov DN, Bozhanova NG, Baranov MS, Soylemez O, Bogatyreva NS *et al.*: **Local fitness landscape of the green fluorescent protein**. *Nature* 2016, **533**:397-401.

23. Lynch M, Conery JS: **The origins of genome complexity**. *Science* 2003, **302**:1401-1404.

24. Weinreich DM, Lan Y, Wylie CS, Heckendorn RB: **Should evolutionary geneticists worry about higher-order epistasis?** *Curr Opin Genet Dev* 2013, **23**:700-707.

25. Podgornaia AI, Laub MT: **Protein evolution. Pervasive
• degeneracy and epistasis in a protein–protein interface**. *Science* 2015, **347**:673-677.
A comprehensive insight into the molecular mechanisms shaping the evolution of functional specificity in protein-protein interactions. The sequence space encompassing the interface formed by PhoQ-PhoP proteins (constituting bacterial two-component signaling system) was mapped via a saturated mutagenesis of four key residues of PhoQ followed by a functional selection. The authors found that degeneracy and epistasis shape the connectivity pattern in sequence space of the identified functional variants.

26. Smith JM: **Natural selection and the concept of a protein space**. *Nature* 1970, **225**:563-564.

27. Wroe R, Chan HS, Bornberg-Bauer E: **A structural model of latent evolutionary potentials underlying neutral networks in proteins**. *HFSP J* 2007, **1**:79-87.

28. van Nimwegen E, Crutchfield JP, Huynen M: **Neutral evolution of mutational robustness**. *Proc Natl Acad Sci U S A* 1999, **96**:9716-9720.

29. Raval A: **Molecular clock on a neutral network**. *Phys Rev Lett* 2007, **99**:138104.

30. Bastolla U, Porto M, Eduardo Roman MH, Vendruscolo MH: **Connectivity of neutral networks, overdispersion, and structural conservation in protein evolution**. *J Mol Evol* 2003, **56**:243-254.

31. Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ: **Empirical fitness landscapes reveal accessible evolutionary paths**. *Nature* 2007, **445**:383-386.

32. Starr TN, Thornton JW: **Epistasis in protein evolution**. *Protein Sci*
• 2016, **25**:1204-1218.
An insightful review that summarizes the current understanding of the prevalence, mechanisms, and evolutionary role of intra-molecular and inter-molecular epistasis.

33. Kaltenbach M, Jackson CJ, Campbell EC, Hollfelder F, Tokuriki N:
•• **Reverse evolution leads to genotypic incompatibility despite functional and active site convergence**. *Elife* 2015:4.
This work uses a directed evolution approach to experimentally test the evolutionary reversibility of an enzyme. Kaltenbach *et al.* evolved phos-photriesrerase to an arylesterase, and back, and found that extensive epistasis makes the reverse evolution genotypically irreversible. But, despite the genotypic entrenchment, phenotypic reversion was possible via fixation of new mutations.

34. Palmer AC, Toprak E, Baym M, Kim S, Veres A, Bershtein S, Kishony R: **Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes**. *Nat Commun* 2015, **6**:7385.

35. Steinberg B, Ostermeier M: **Environmental changes bridge evolutionary valleys**. *Sci Adv* 2016, **2**:e1500921.

36. Tian J, Woodard JC, Whitney A, Shakhnovich EI: **Thermal stabilization of dihydrofolate reductase using Monte Carlo unfolding simulations and its functional consequences**. *PLoS Comput Biol* 2015, **11**:e1004207.

37. Bridgham JT, Ortlund EA, Thornton JW: **An epistatic ratchet constrains the direction of glucocorticoid receptor evolution**. *Nature* 2009, **461**:515-519.

38. Anderson DW, McKeown AN, Thornton JW: **Intermolecular**
•• **epistasis shaped the function and evolution of an ancient transcription factor and its DNA binding sites**. *Elife* 2015, **4**:e07864.
An important work that provides a comprehensive insight into the role of intra-molecular and inter-molecular epistasis in the evolution of molecular interfaces. The joint sequence space connecting the evolutionary transition in specificity of a transcription factor (TF) and a corresponding DNA response element (RE) was experimentally mapped, and it was concluded that epistasis was essential for the evolution of TF–RE complexes with unique specificity.

39. Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW: **Crystal structure of an ancient protein: evolution by conformational epistasis**. *Science* 2007, **317**:1544-1548.

40. Natarajan C, Inoguchi N, Weber RE, Fago A, Moriyama H, Storz JF: **Epistasis among adaptive mutations in deer mouse hemoglobin**. *Science* 2013, **340**:1324-1327.

41. Gong LI, Suchard MA, Bloom JD: **Stability-mediated epistasis**
•• **constrains the evolution of an influenza protein**. *Elife* 2013, **2**:e00631.
This work identifies protein thermodynamic stability as the molecular mechanism of the epistatic events accompanying the evolution of influenza nucleoprotein (NP). By reconstructing a 39 mutation long evolutionary trajectory of the NP and testing the effects of these mutations independently on a parent background, Gong *et al.* found that several mutations were destabilizing and were preceded by acquisition of stabilizing mutations that buffered their otherwise deleterious effects.

42. Olson CA, Wu NC, Sun R: **A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain**. *Curr Biol* 2014, **24**:2643-2651.

43. Pollock DD, Thiltgen G, Goldstein RA: **Amino acid coevolution induces an evolutionary stokes shift**. *Proc Natl Acad Sci U S A* 2012, **109**:E1352-E1359.

44. Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS: **The stability effects of protein mutations appear to be universally distributed**. *J Mol Biol* 2007, **369**:1318-1332.

45. Tokuriki N, Stricher F, Serrano L, Tawfik DS: **How protein stability and new functions trade off**. *PLoS Comput Biol* 2008, **4**:e1000002.

46. DePristo MA, Weinreich DM, Hartl DL: **Missense meanderings in sequence space: a biophysical view of protein evolution**. *Nat Rev Genet* 2005, **6**:678-687.

47. Studer RA, Christin PA, Williams MA, Orengo CA: **Stability-activity tradeoffs constrain the adaptive evolution of rubisco**. *Proc Natl Acad Sci U S A* 2014.

48. Beadle BM, Shoichet BK: **Structural bases of stability-function tradeoffs in enzymes**. *J Mol Biol* 2002, **321**:285-296.

49. Bloom JD, Labthavikul ST, Otey CR, Arnold FH: **Protein stability promotes evolvability**. *Proc Natl Acad Sci U S A* 2006, **103**:5869-5874.

50. Zeldovich KB, Chen P, Shakhnovich EI: **Protein stability imposes limits on organism complexity and speed of molecular evolution**. *Proc Natl Acad Sci U S A* 2007, **104**:16152-16157.

51. Olsson MH, Parson WW, Warshel A: **Dynamical contributions to enzyme catalysis: critical tests of a popular hypothesis**. *Chem Rev* 2006, **106**:1737-1756.

52. Haq O, Andrec M, Morozov AV, Levy RM: **Correlated electrostatic mutations provide a reservoir of stability in HIV protease**. *PLoS Comput Biol* 2012, **8**:e1002675.

53. Elena SF, Lenski RE: **Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation**. *Nat Rev Genet* 2003, **4**:457-469.

54. Lenski RE, Bennett AF: **Evolutionary response of escherichia coli to thermal stress**. *Am Nat* 1993, **142(Suppl. 1)**:S47-S64.

55. Lunzer M, Golding GB, Dean AM: **Pervasive cryptic epistasis in molecular evolution**. *PLoS Genet* 2010, **6**:e1001162.

56. Breen MS, Kemena C, Vlasov PK, Notredame C, Kondrashov FA: **Epistasis as the primary factor in molecular evolution**. *Nature* 2012, **490**:535-538.

57. Ashenberg O, Gong LI, Bloom JD: **Mutational effects on stability are largely conserved during protein evolution**. *Proc Natl Acad Sci U S A* 2013, **110**:21071-21076.

58. Doud MB, Ashenberg O, Bloom JD: **Site-specific amino acid preferences are mostly conserved in two closely related protein homologs**. *Mol Biol Evol* 2015, **32**:2944-2960.

59. Pal C, Papp B, Lercher MJ: **An integrated view of protein evolution**. *Nat Rev Genet* 2006, **7**:337-348.

60. Fares MA, Barrio E, Sabater-Munoz B, Moya A: **The evolution of the heat-shock protein groel from buchnera, the primary endosymbiont of aphids, is governed by positive selection**. *Mol Biol Evol* 2002, **19**:1162-1170.

61. Maisnier-Patin S, Roth JR, Fredriksson A, Nystrom T, Berg OG, Andersson DI: **Genomic buffering mitigates the effects of deleterious mutations in bacteria**. *Nat Genet* 2005, **37**:1376-1379.

62. Oliver A, Canton R, Campo P, Baquero F, Blazquez J: **High frequency of hypermutable pseudomonas aeruginosa in cystic fibrosis lung infection**. *Science* 2000, **288**:1251-1254.

63. Aguilar-Rodriguez J, Sabater-Munoz B, Montagud-Martinez R, Berlanga V, Alvarez-Ponce D, Wagner A, Fares MA: **The molecular chaperone dnak is a source of mutational robustness**. *Genome Biol Evol* 2016.

64. Queitsch C, Sangster TA, Lindquist S: **Hsp90 as a capacitor of phenotypic variation**. *Nature* 2002, **417**:618-624.

65. Rohner N, Jarosz DF, Kowalko JE, Yoshizawa M, Jeffery WR, Borowsky RL, Lindquist S, Tabin CJ: **Cryptic variation in morphological evolution: Hsp90 as a capacitor for loss of eyes in cavefish**. *Science* 2013, **342**:1372-1375.

66. Rutherford SL, Lindquist S: **Hsp90 as a capacitor for morphological evolution**. *Nature* 1998, **396**:336-342.

67. Cho Y, Zhang X, Pobre KF, Liu Y, Powers DL, Kelly JW, Gierasch LM, Powers ET: **Individual and collective contributions of chaperoning and degradation to protein homeostasis in *E. coli***. *Cell Rep* 2015, **11**:321-333.

68. Bogumil D, Dagan T: **Chaperonin-dependent accelerated substitution rates in prokaryotes**. *Genome Biol Evol* 2010, **2**:602-608.

69. Lachowiec J, Lemus T, Borenstein E, Queitsch C: **Hsp90 promotes kinase evolution**. *Mol Biol Evol* 2015, **32**:91-99.

70. Tokuriki N, Tawfik DS: **Chaperonin overexpression promotes genetic variation and enzyme evolution**. *Nature* 2009, **459**:668-673.

71. Cetinbas M, Shakhnovich EI: **Catalysis of protein folding by chaperones accelerates evolutionary dynamics in adapting cell populations**. *PLoS Comput Biol* 2013, **9**:e1003269.

72. Kadibalban AS, Bogumil D, Landan G, Dagan T: **Dnak-dependent accelerated evolutionary rate in prokaryotes**. *Genome Biol Evol* 2016, **8**:1590-1599.

73. Bershtein S, Serohijos AW, Bhattacharyya S, Manhart M, Choi JM,
•• Mu W, Zhou J, Shakhnovich EI: **Protein homeostasis imposes a barrier on functional integration of horizontally transferred genes in bacteria**. *PLoS Genet* 2015, **11**:e1005612.
The authors created 35 new *E. coli* strains where endogenous folA gene (encoding DHFR) had been replaced on the chromosome by the orthologous genes from other mesophilic bacteria of various degree of divergence from *E. coli*. A concomitant determination of all the molecular properties of orthologous DHFRs and their intracellular abundances in *E. coli* cytoplasm allowed to establish a quantitative biophysical fitness landscape for DHFR and highlighted the role of PQC in shaping fitness effects of horizontal transfer of orthologues. This approach made it

possible to determine the effect of variation of biophysical parameters of an essential enzyme on fitness in a broad range of variation that is not achievable by exploration of point mutations only.

74. Dykhuizen DE, Dean AM, Hartl DL: **Metabolic flux and fitness**. *Genetics* 1987, **115**:25-31.

75. Kacser H, Burns JA: **The molecular basis of dominance**. *Genetics* 1981, **97**:639-666.

76. Chen P, Shakhnovich EI: **Lethal mutagenesis in viruses and bacteria**. *Genetics* 2009, **183**:639-650.

77. de Godoy LM, Olsen JV, Cox J, Nielsen ML, Hubner NC, Frohlich F, Walther TC, Mann M: **Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast**. *Nature* 2008, **455**:1251-1254.

78. Zhou L, Zhang AB, Wang R, Marcotte EM, Vogel C: **The proteomic response to mutants of the *Escherichia coli* RNA degradosome**. *Mol Biosyst* 2013, **9**:750-757.

79. Serohijos AW, Shakhnovich EI: **Contribution of selection for protein folding stability in shaping the patterns of polymorphisms in coding regions**. *Mol Biol Evol* 2014, **31**:165-176.

80. Goldstein RA: **Population size dependence of fitness effect distribution and substitution rate probed by biophysical model of protein thermostability**. *Genome Biol Evol* 2013, **5**:1584-1593.

81. Crow JF, Kimura M: *An Introduction to Population Genetics Theory*. New York: Harper & Row; 1970.

82. Shakhnovich EI, Finkelstein AV: **Theory of cooperative transitions in protein molecules. 1. Why denaturation of globular protein is a 1st-order phase-transition**. *Biopolymers* 1989, **28**:1667-1680.

83. Bloom JD, Silberg JJ, Wilke CO, Drummond DA, Adami C, Arnold FH: **Thermodynamic prediction of protein neutrality**. *Proc Natl Acad Sci U S A* 2005, **102**:606-611.

84. Dokholyan NV, Shakhnovich EI: **Understanding hierarchical protein evolution from first principles**. *J Mol Biol* 2001, **312**:289-307.

85. Taverna DM, Goldstein RA: **Why are proteins so robust to site mutations?** *J Mol Biol* 2002, **315**:479-484.

86. Privalov PL, Khechinashvili NN: **A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study**. *J Mol Biol* 1974, **86**:665-684.

87. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L: **The foldx web server: an online force field**. *Nucleic Acids Res* 2005, **33(Web Server issue)**:W382-W388.

88. Yin S, Ding F, Dokholyan NV: **Eris: an automated estimator of protein stability**. *Nat Methods* 2007, **4**:466-467.

89. Miyazawa S, Jernigan RL: **Estimation of effective interresidue contact energies from protein crystal-structures — quasi-chemical approximation**. *Macromolecules* 1985, **18**:534-552.

90. Sarai A, Gromiha MM, An JH, Prabakaran P, Selvaraj S, Kono H, Oobatake M, Uedaira H: **Thermodynamic databases for proteins and protein–nucleic acid interactions**. *Biopolymers* 2001, **61**:121-126.

91. Fersht AR, Matouschek A, Serrano L: **The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding**. *J Mol Biol* 1992, **224**:771-782.

92. Kimura M: **On the probability of fixation of mutant genes in a population**. *Genetics* 1962, **47**:713-719.

93. Dasmeh P, Serohijos AW, Kepp KP, Shakhnovich EI: **The influence of selection for protein stability on dn/ds estimations**. *Genome Biol Evol* 2014, **6**:2956-2967.

94. Serohijos AW, Rimas Z, Shakhnovich EI: **Protein biophysics explains why highly abundant proteins evolve slowly**. *Cell Rep* 2012, **2**:249-256.

95. Serohijos AW, Lee SY, Shakhnovich EI: **Highly abundant proteins favor more stable 3d structures in yeast**. *Biophys J* 2013, **104**:L1-L3.

96. Drummond DA, Wilke CO: **Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution**. *Cell* 2008, **134**:341-352.

97. Yang JR, Zhuang SM, Zhang J: **Impact of translational error-induced and error-free misfolding on the rate of protein evolution**. *Mol Syst Biol* 2010, **6**:421.

98. Lobkovsky AE, Wolf YI, Koonin EV: **Universal distribution of protein evolution rates as a consequence of protein folding physics**. *Proc Natl Acad Sci U S A* 2010, **107**:2983-2988.

99. Bank C, Hietpas RT, Jensen JD, Bolon DN: **A systematic survey of an intragenic epistatic landscape**. *Mol Biol Evol* 2015, **32**:229-238.

100. Steinberg B, Ostermeier M: **Shifting fitness and epistatic landscapes reflect trade-offs along an evolutionary pathway**. *J Mol Biol* 2016, **428**:2730-2743.

101. Cheron N, Serohijos AW, Choi JM, Shakhnovich EI: **Evolutionary dynamics of viral escape under antibodies stress: a biophysical model**. *Protein Sci* 2016, **25**:1332-1340.

102. Dixit PD, Maslov S: **Evolutionary capacitance and control of protein stability in protein–protein interaction networks**. *PLoS Comput Biol* 2013, **9**:e1003023.

103. Chen P, Shakhnovich EI: **Thermal adaptation of viruses and bacteria**. *Biophys J* 2010, **98**:1109-1118.