REVIEW ARTICLE
DOI: 10.1038/s41559-017-0228-1

nature ecology & evolution

Evolution of complex adaptations in molecular systems

Csaba Pál* and Balázs Papp

A central challenge in evolutionary biology concerns the mechanisms by which complex adaptations arise. Such adaptations depend on the fixation of multiple, highly specific mutations, where intermediate stages of evolution seemingly provide little or no benefit. It is generally assumed that the establishment of complex adaptations is very slow in nature, as evolution of such traits demands special population genetic or environmental circumstances. However, blueprints of complex adaptations in molecular systems are pervasive, indicating that they can readily evolve. We discuss the prospects and limitations of non-adaptive scenarios, which assume numerous neutral or deleterious steps in the evolution of complex adaptations. Next, we examine how complex adaptations can evolve by natural selection in a changing environment. Finally, we argue that molecular 'springboards' such as phenotypic heterogeneity and promiscuous interactions facilitate this process by providing access to new adaptive paths.

What are the limits to perfection in nature[1]? Certain molecular traits may be difficult to evolve, not because they are made impossible by laws of physics and chemistry, but because multiple mutations must be fixed together to provide a benefit (Fig. 1). Here, such traits will be referred to as complex adaptations. Darwin himself was highly aware of this problem[2]. He stated that, "if it could be demonstrated that any complex organ existed, which could not possibly have been formed by numerous, successive, slight modifications, my theory would absolutely break down". Complex adaptations highlight a persistent problem in evolutionary discussions. Many analyses focus on only the adaptive nature of the end result of evolution. However, what looks like an optimal organism design may not be attainable by a series of adaptive mutations.

The problem of complex adaptation relates to another central issue in evolutionary biology. In the early 1920s, Ronald Fisher pioneered the view that adaptation is by and large a hill-climbing process: it proceeds through progressive accumulation of beneficial mutations[3]. In contrast, Sewall Wright proposed that fixation of conditionally deleterious mutations acts as stepping stones by providing access to evolutionary routes that are otherwise inaccessible[4]. This issue has practical implications. For example, directed protein evolution typically employs random mutagenesis coupled with strong functional selection[5]. Such protocols generally fail to discover new enzymatic activities that demand the accumulation of many neutral or slightly deleterious mutations. Computer scientists face similar problems: standard evolutionary algorithms have a tendency to quickly converge on a local solution and hence frequently fail to identify more promising regions of the parameter space[6].

There are two fundamental problems with reconstructing the evolutionary history of complex adaptations: shortage of intermediate forms in extant organisms and improbable fossilization ('missing links'). This Review focuses on complex adaptations in cellular systems, where these problems are especially severe. Recent advances in experimental evolution, deep-scan mutational assays and systems biology have allowed us to study the problem of complex adaptation in a rigorous manner. These studies unequivocally demonstrated that complex adaptations in molecular systems are pervasive and can readily evolve in the laboratory.

It is worth emphasizing what this Review is not about. We do not discuss the concept of facilitated variation[7,8], modularity[9] or the evolution of biological complexity per se, as these topics are not directly related and have been reviewed elsewhere. We do not explicitly review key innovations, defined as phenotypic traits that allow subsequent evolutionary radiation of a taxonomic group[10]. Evolution of key innovations could be limited by physicochemical constraints, environmental conditions or shortage of mutational combinations required for functional change. Therefore, the notions of key innovations and complex adaptation are different from each other.

We start by providing a classification scheme with examples, followed by prospects and limitations of current theories on how and why complex adaptations arise. We introduce conceptual (Box 1) and methodological advances (Box 2), and end the paper by discussing future directions and predictions that have so far remained untested.

Classification of complex adaptations

Complex adaptations can occur within a single gene (intramolecular; Fig. 2). One form involves mutually dependent mutations that stabilize local elements in a protein structure and simultaneously promote new functions[11]. Here, researchers face the 'chicken or the egg' dilemma: mutations in the enzymatic active site are typically destabilizing, whereas structural mutations elsewhere in the protein provide no obvious change in enzymatic function but mitigate the fitness cost of the former. It is not trivial which of the two types of mutation comes first. Another common form of intramolecular complex adaptation concerns the establishment of pairwise interactions between interdependent sites within a single RNA or protein molecule. Well-known examples include the origin of stem–loop interactions in transfer RNA[12] and disulfide bonds in protein molecules[13]. As the disruption of such highly specific interactions often compromises stability or function, explaining their evolution as a stepwise process is challenging.

Intermolecular complex adaptations involve genetic changes at numerous loci and belong to three main classes (Fig. 2). The most elementary form entails the establishment of new interactions between two macromolecules, such as seen in transcription

Synthetic and Systems Biology Unit, Biological Research Center, Szeged 6726, Hungary. *e-mail: cpal@brc.hu

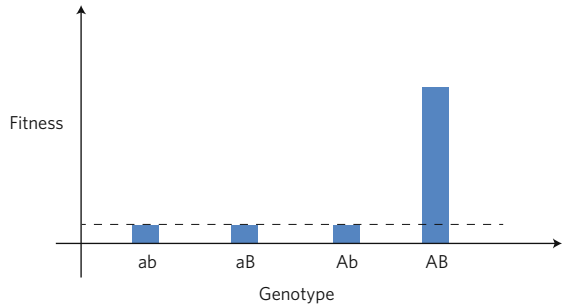

Figure 1 | Concept of complex adaptations. Two mutations (a→A and b→B) have to occur simultaneously to provide a fitness advantage. Note that the individual mutations depicted here are neutral, but they could also be deleterious.

factor–binding site interactions, two-component signal transduction systems and, more generally, protein–protein binding interfaces. Although complementarity is essential for maintenance of function, these systems frequently show substantial divergence across species[14]. The main issue is how new interaction specificities arise without deleterious cross-talks or non-functional intermediate stages. The second class involves the origin of multi-gene molecular pathways that demand the coordinated operation of multiple gene products. A prime example is metabolic pathways that typically work through a series of enzymatic steps to produce a desired end-product[15]. As the intermediate steps do not necessarily produce useful metabolites, explaining the stepwise evolution of such pathways demands careful consideration. The third class of complex adaptations concerns the origin of macromolecular cellular structures, such as bacterial flagella[16,17], eukaryotic centrioles and ion channels. Such structures require the establishment of new specific interaction events and the coordinated action of numerous proteins. Although precursors of these complex traits are widespread across species, deciphering step-by-step evolutionary histories has proved to be generally difficult. Multimeric protein complexes are generally assembled from subunits derived from gene duplication[16,18], but the evolutionary forces driving their diversification remain largely unknown.

Leaps in the genotype space

In his *New Essays on Human Understanding*, Leibniz proposed the principle of continuity in nature. This maxim states that nature does not make jumps ("*natura non facit saltum*"): objects and properties in nature change gradually, rather than suddenly. Darwin was also very explicit about this issue[2]. He stated that natural selection proceeds by the accumulation of numerous slightly favourable steps, all of which have individually small effects on the performance of organisms. Similarly, modern scholars generally state that evolution proceeds in small discrete steps in the genotype space, where steps are limited to immediate mutational neighbours (but see refs 19,20). The justification for this assumption is that the rise of double mutants is usually negligible, and mutations with large effects tend to have deleterious side consequences. We do not discuss the validity of these assumptions in detail, but rather focus on some notable exceptions.

Several mutational mechanisms initiate major phenotypic transitions. First, a sizeable fraction of nucleotide substitutions occur simultaneously within short stretches of DNA[21] and may therefore affect multiple interacting sites within a protein[22]. Moreover, structural alterations such as deletions and insertions induce point mutations in nearby regions[23,24]. Second, as recombination events can reorganize protein modules, proteins or entire cellular networks, they can lead to evolutionary novelties inaccessible by only point mutations[25,26]. Third, and on a related note, horizontal gene transfer—a major source of evolutionary novelties in bacteria—often results in the simultaneous acquisition of numerous functionally related genes[27,28]. Fourth, large-scale duplications and aneuploidy events simultaneously increase the dosage of multiple genes, and thereby produce large phenotypic leaps[29]. A case study on adaptive evolution in yeast demonstrated that extra copies of a chromosome increased fitness by simultaneous overexpression of two specific regulatory genes[30]. Remarkably, overexpressing any of the two genes individually provided no fitness advantage[30]. The frequency and mechanisms by which gross chromosomal mutations aid the crossing of suboptimal intermediate states remains an important open question for future studies.

Non-adaptive origins of complex adaptations

How do complex adaptations arise in multiple steps? One theory suggests that they involve mutations that are individually neutral or even deleterious, but together provide a fitness advantage[31–33]. Non-adaptive mutations may accumulate by genetic drift first, and thereby prepare the ground for later mutations that confer new, adaptive traits. Alternatively, multiple mutations may be fixed simultaneously in large populations (see Fig. 3a and Box 1 for more details).

The best evidence for the theory comes from laboratory evolution studies on catalytic RNAs[34] and proteins[5,32,35]. Several works demonstrated that neutral mutations promote the evolution of new protein functions in the laboratory[35,36]. Moreover, large populations of RNA enzymes with accumulated neutral variation adapted more rapidly to cleave a new RNA substrate than a population without such genetic variation[34].

Non-adaptive evolution may also contribute to the initial establishment of altered multi-protein complexes[37,38] and DNA regulatory interactions[39] prior to adaptive diversification. For example, epistasis across the molecular interface of a transcription factor and the binding site allows the evolution of new binding specificities to arise by neutral evolution[39]. Similarly, a new subunit in a fungal proton pump arose by duplication followed by complementary degeneration of binding interfaces in the two copies. During this process, both copies became obligate components of the complex without conferring any new function[40].

Despite its apparent success, the theory faces several problems. First, when high-fitness genotypes are well isolated from each other and sparse on the adaptive landscape, molecules subjected to genetic drift will fail to escape the local fitness peak. Indeed, the protocol of neutral drift followed by positive selection has met with mixed success in certain studies[41,42]. For instance, drift produced only protein variants that had also been described in conventional directed evolution experiments based on only positive selection[42].

Second, laboratory studies of genetic drift in protein and RNA molecules typically applied exceedingly high mutation rates[32,34,35]. Whether neutral mutations promote the evolution of new molecular functions under more realistic population genetic settings remains to be tested.

Third, the role of non-adaptive forces is questionable in the origin of multi-step pathways. Most notably, it has been proposed that "additions of individual reactions to a metabolic network will not change the phenotype until a second added reaction connects the first one to an already existing pathway"[33]. There is no direct empirical support for this scenario in bacteria, which are especially prolific in producing metabolic innovations[43]. Genes under no selection are rapidly inactivated and subsequently lost in free-living bacteria, not least because there is a pervasive mutational bias towards deletions of genomic segments[44,45].

Finally, numerous non-adaptive steps may render the establishment of new complex traits exceedingly unlikely. The process is

Box 1 | Complex adaptations—theoretical considerations

By definition, complex adaptations are phenotypic traits requiring multiple, specific mutations to yield a functional advantage. Despite substantial efforts, the population genetic mechanisms driving complex adaptations are unclear[99,100]. In a nutshell, the paradox is as follows. As mutational events are rare, complex adaptations are unlikely to occur in small populations: the waiting time for the rise of multiple specific mutations would be very long. This would suggest that the evolution of complex adaptations is facilitated in large populations. The issue is more complicated, though. If the intermediate mutational steps towards complex adaptation are individually deleterious, they will be purged in large populations. As the fixation of deleterious mutations is exceedingly unlikely with growing population size, shift from one adaptive peak to another through weakly deleterious intermediates can occur only when genetic drift prevails, that is, in small populations.

How can the seeming paradox be resolved? First, the intermediate mutation may confer a benefit under alternative environmental or genetic conditions (see main text). Alternatively, we may need to abandon the idea of sequential fixation of intermediate mutations. Although individual deleterious intermediate-stage mutations have negligible chance of fixation in large populations, there is a steady input of such mutations. Indeed, segregating deleterious mutations are common in natural populations of yeast[101] and in humans[102,103]. Such a stable reservoir of non-adaptive mutations is poised for the rise of a second mutation that is adaptive in the specific genetic background, leading to simultaneous fixation of the two mutations. As the size of the reservoir increases with growing population size, the time required for this process declines sharply as population size increases[88,104,105]. Thus, selection alone can offer a solution to escape local fitness peaks in natural populations. One may argue that this theory is unlikely to work in the case of three or more non-adaptive intermediate steps, as the simultaneous rise of multiple mutations in a single genotype is exceedingly unlikely. Although this criticism may hold when the intermediate steps are deleterious, the effect is partly offset by the elevated number of paths towards the final adaptation when the intermediate states are neutral[105].

Evolutionary escape from local fitness peaks has been a central problem for more than 85 years now[99]. Unfortunately, the lack of consensus on the relative roles played by mutation, recombination and random genetic drift hinders empirical tests. The role of recombination is particularly controversial: it can either facilitate the escape from local fitness peaks by combining mutations from different individuals, or hinder it by breaking up adaptive combinations[106]. As a result, only low recombination rates can speed up the crossing of fitness valleys, while high rates are predicted to be disadvantageous[106].

Rates of recombination and mutation, and the importance of genetic drift vary enormously across evolutionary lineages, but do they influence which evolutionary pathways are realized in nature? For instance, it is currently unknown whether multicellular eukaryotic species with relatively low population size and high mutation rate have greater or smaller acquisition of complex adaptations on a per generation basis[105]. As a further complication, the answer probably depends on the molecular basis of adaptation, for example, the number of sites involved and whether the intermediate states are deleterious or neutral.

Finally, most theoretical considerations have focused on the rate of traversing a single specific mutational path, which is expected to be low when neutral or deleterious mutations are involved. However, in realistic fitness landscapes there could be numerous different possible mutational paths to the same genotype, and there might be several possible beneficial genotypes that are phenotypically equivalent. As a result, evolution may follow trajectories that involve neutral or deleterious intermediates even in the presence of directly uphill trajectories[106]. Addressing this issue demands quantification of the frequencies of these different modes of mutational trajectories on realistic fitness landscapes.

expected to be much faster with the availability of adaptive bypasses (see below). Given these considerations, it is important to consider scenarios that assume stepwise evolution of complex adaptations by natural selection.

Preadaptations and indirect evolutionary paths

As Darwin himself elaborated, many complex traits evolved from earlier traits that had served different functions[2]. Examples for molecular preadaptations are abundant in the literature. For instance, several components of the bacterial flagellum share homologous proteins with the type III secretion system, indicating that one evolved from the other[17]. Another example is the citrate acid cycle: it most probably originated by assembling chemical steps previously functioning in amino acid biosynthesis[46,47]. Similarly, studies on digital organisms revealed that populations often evolve complex features by building on simpler functions that had evolved earlier[48]. Accordingly, early rising mutations serve as stepping stones in the evolution of complex traits.

The theory has three main predictions. First, asymmetric relationships between proteins or cellular subsystems should be common: the function of one protein (A) depends on another protein (B), but the function of protein B does not depend on A. The best examples come from metabolism[28,49]. When multiple enzymes produce the same metabolite, enzymes in the converging reaction (A) depend on the metabolite flux through B, but not vice versa (Fig. 3b). Such asymmetry is reflected in gene expression, gene essentiality and most importantly in the evolution of gene content[49]. Similarly, the functional relationships between regulators and target proteins are also frequently asymmetric: the target may exert its function without the regulator in specific contexts, but not necessarily vice versa. Indeed, transcription factors in bacteria are typically less conserved than their target genes and evolve independently of them[50]. More generally, large-scale genetic interaction screens uncovered a plethora of asymmetric functional relationships between gene pairs where mutation in one gene modifies the effect of mutation in the other[51].

Second, as a consequence of asymmetrical protein relationships, evolution should proceed via a defined order[43,49]. In agreement with the expectations, enzymes that serve as stepping stones towards multi-step adaptations were gained on an earlier branch of the phylogenetic tree compared with enzymes that provide benefit only in the presence of the partner enzymes[43]. More generally, genes pre-existing in an organism should influence the functionality of a horizontally acquired gene product if it operates on an ancestral pathway[52]. A recent large-scale analysis of bacterial genomes provided general support for this idea[53]. It showed that specific molecular functions tend to be gained sequentially by horizontal gene transfer[53], suggesting that bacterial evolution is governed by functional assembly patterns. Other studies aimed to reconstruct adaptive landscapes in a single protein by functional analysis of mutational combinations[54]. It seems that evolutionary novelties by horizontal gene transfer and point mutations are profoundly constrained by epistasis, and therefore frequently occur via a defined order.

Box 2 | Methods to study complex adaptations

Population genetics has a long tradition of studying the problem of complex adaptations[4]. Theoretical studies indicate that the time taken to establish complex adaptations depends on population size, mutation rate, recombination and the magnitude of the selective disadvantage of intermediate-state alleles (Box 1). Specific examples of complex adaptations can be studied by phylogenetic analysis, molecular laboratory experiments and computational systems biology.

Functionally related genes that are jointly needed for a cellular function are not expected to be accessible gradually, and therefore they should arise and be lost together on the same branch of the phylogenetic tree. Comparative genomics has provided ample evidence for such co-evolving gene clusters, especially in the case of macromolecular protein complexes[107] and linear metabolic pathway[28]. Other complex traits are assumed to be accessible only in a defined order[49,53], as certain mutations pave the way for subsequent adaptive changes.

Molecular studies in the laboratory have unequivocally demonstrated that complex adaptations in molecular systems are pervasive. Three main types of analysis have been deployed: mutational analysis coupled with exploration of the adaptive landscape, reconstruction of ancestral molecules and laboratory experimental evolution. Deep-scan mutational scanning showed that in spite of the prevalence of harmful mutations, mutational effects vary due to epistatic interactions with other mutations[108]. Reconstruction and functional analysis of ancestral molecules showed that adaptive evolution frequently demands prior fixation of other, functionally silent mutations[89,109]. Directed protein evolution studies investigated the impact of neutral exploration of the genotype space on acquisition of new enzymatic functions (see main text for details).

Finally, systems biology models offer a new angle to study the underlying molecular processes of complex adaptations that involve numerous gene products[110]. By providing a mapping between genotype and phenotype, molecular network models provide valuable insights into the exploration of the adaptive landscape, with the ultimate goal of predicting which particular evolutionary trajectories are realized, while others are not. For instance, using genome-scale metabolic modelling, recent works tested the evolutionary mechanisms whereby complex bacterial metabolic innovations arise[43,110]. In the main text, we report on the use of these methodological developments for testing theories of complex adaptations.

Third, evolutionary traps can be circumvented by 'extra-dimensional bypasses' (Fig. 3c). Accordingly[55], the necessity of non-adaptive steps reflects limited dimensionality of the adaptive landscapes considered, and therefore is more apparent than real (Box 1). Indeed, most molecular evolution studies are confined to the immediate neighbourhood around the wild-type sequence and are limited in scope. A recent study investigated this problem systematically[55]. It confirmed that many direct evolutionary paths are indeed blocked by pairwise epistatic interactions (that is, populations along these paths must proceed via one or more non-adaptive steps). By exhaustive analysis of 160,000 mutational variants of a single protein, the authors show that such evolutionary traps can be circumvented by indirect paths through gain and subsequent loss of mutations[55,56]. These results indicate that higher-order epistasis is critically important for understanding evolutionary processes[57].

Dynamic environment

The dynamic environment scenario heavily relies on the notion of preadaptation[43,58,59]. It claims that stepping-stone mutations fixed earlier are beneficial in only specific environments and therefore complex adaptation is accelerated in varying environments (Fig. 3d). Computer simulations of genetic circuits and RNA molecules[60], and verbal arguments on early expansion of molecular pathways[61], reached similar conclusions.

The scenario is attractive for two reasons. Environmental change is ubiquitous and epistasis changes qualitatively across environments[62–64]. Mutations that are neutral or even deleterious in one environment may turn beneficial in another environmental context. Therefore, environmental change may facilitate evolution by exposing populations to new adaptive routes. Recent studies provide direct empirical support for the theory. The examples include

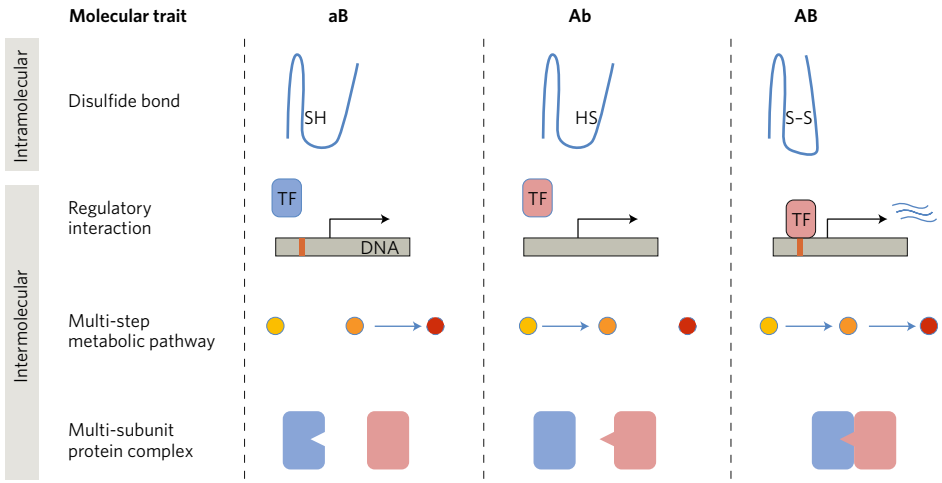

Figure 2 | Main classes and examples of complex adaptations in molecular systems. Establishment of a new disulfide bond (S–S) from two adjacent sulfhydryl groups (–SH) within the same protein molecule represents an example of intramolecular complex adaptation. The origin of new transcription factor–DNA binding site interactions, multi-step metabolic pathways and multi-subunit complexes all qualify as intermolecular complex adaptations requiring specific mutations in multiple genes. Yellow, orange and red circles represent a metabolic pathway's substrate, intermediate metabolite and end product, respectively.

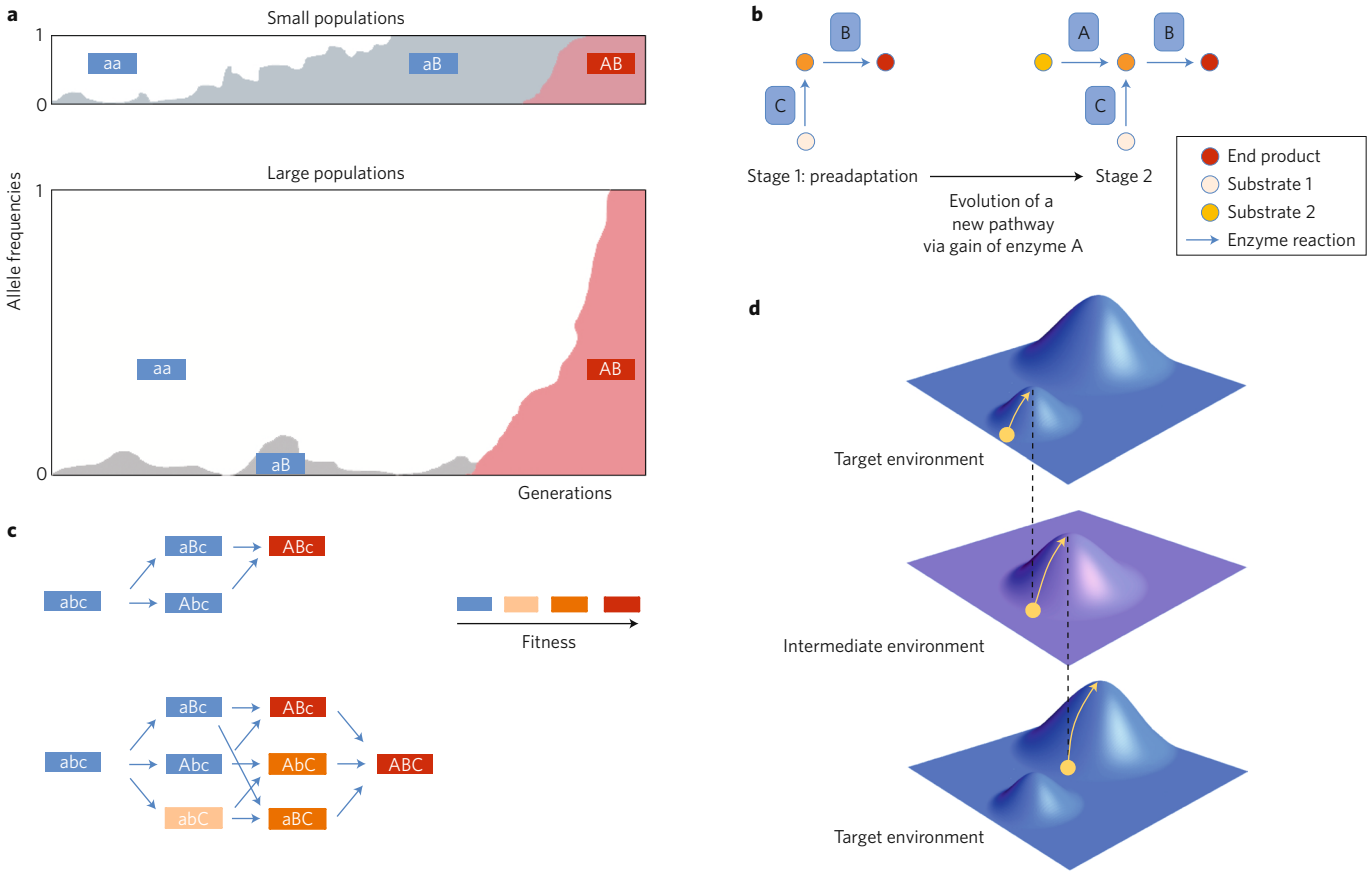

Figure 3 | Evolutionary mechanisms of establishing complex adaptations. a, Non-adaptive origin of complex adaptations can occur through sequential fixation of mutations in small populations where the intermediate mutation first goes to fixation through genetic drift (upper panel). Alternatively, in large populations a beneficial second mutation can arise in a descendant of the intermediate mutation and they fix simultaneously without the population ever going through a pure intermediate state (lower panel). Allele frequency plots depict the dynamics of the neutral intermediate mutation and the beneficial second mutation. **b**, Simplified metabolic network scheme in which enzyme functions A and B are asymmetrically dependent on each other as a consequence of preadaptation. In the first evolutionary stage, the network can metabolize substrate 1 and this serves as a preadaptation to use substrate 2 via the acquisition of enzyme A. As in the second stage both enzymes A and C produce the same intermediate metabolite, the activity of downstream enzyme B does not exclusively depend on A. In contrast, A can be active under steady state conditions only when B is active, hence their functional dependence is asymmetric. **c**, Evolution of complex adaptations via adaptive by-passes in extra genotype dimensions. The figure depicts a simplified scenario where evolution from a low-fitness genotype (ab) to a high-fitness one (AB) involves neutral intermediary steps, but a mutation at a third locus (c→C) opens new uphill trajectories where all intermediate steps are beneficial. Fitness of genotypes is represented by a colour scale. **d**, Evolution in alternating environments promotes escape from local fitness optima through a series of purely adaptive walks. The horizontal plane represents genotype space, the vertical axis represents fitness and arrows indicate uphill evolutionary trajectories. Environmental change alters the fitness landscape in such a way that a fitness valley in the target environment becomes a fitness peak in the intermediate environment, hence facilitating valley crossing.

the evolution of new protein functions, transcription factor–DNA binding site interfaces and metabolic pathways.

Protein evolution was examined by analysing single amino acid mutants in an enzyme under selection for a wild-type function and for a new function[65,66]. Interestingly, resistance to a new antibiotic emerged from mutations that are neutral to another drug at low levels but deleterious at high levels; thus the capacity to evolve a new function also depends on the strength of selection[65,66]. It seems that fluctuating environments select for enzymes with especially high activities[66]. Another study explicitly demonstrated that alteration of the adaptive landscape by environmental change permits exploration of new regions of the sequence space that are otherwise selected against, and this process leads to superior phenotypes[67].

The lac operon in *Escherichia coli* tells a similar story[68]. In a fixed environment, the evolution of transcription factor and binding sites is trapped at a local solution, as beneficial mutational combinations cannot be reached by only natural selection. However, in a fluctuating environment, many of them become accessible via adaptive steps. Such escapes from adaptive stasis readily occurred because numerous mutations had opposite effects in the two environments and therefore environmental changes opened up new possibilities for beneficial mutations[68].

Finally, it has been proposed that temporally varying nutrient conditions selects for new enzymatic reactions that, as a by-product, serve as stepping stones towards the establishment of complex metabolic pathways[43]. Three complementary approaches provided support for the theory. First, computational analysis of bacterial metabolic networks revealed that new complex pathways can evolve via the acquisition of single biochemical reactions that confer a benefit under specific environmental conditions[43]. Second, by reconstructing the evolutionary history of gene gains in bacteria, it was shown that complex metabolic pathways are indeed established in a defined order as predicted by the dynamic environment model[43]. Third, laboratory evolution studies showed that adaptation to one carbon source promotes the utilization of other nutrients[43,69].

Together, these studies indicate that complex traits can emerge in complex environments without the need to invoke neutral exploration of genotype space, a view that is in sharp contrast with non-adaptive scenarios of complex adaptations. This conclusion has important ramifications for those studying the design principles of complex molecular pathways and those aiming to create industrially useful microbes. First, deciphering the adaptive value of molecular pathways might often require studying their operation under numerous environmental conditions. Second, evolutionary engineering of microbes to obtain desired phenotypes may demand complex, temporally varying selection pressure in the laboratory.

Although the dynamic environment model is a promising alternative to non-adaptive scenarios, it nevertheless faces several conceptual challenges. Most notably, it is unclear why mutations that are adaptive in one transient environment are not selected against and lost upon environmental change. A major barrier to complex adaptations may be the absence of a relevant series of environmental conditions, rather than the shortage of relevant mutations in the population. Answering this question demands the network of evolutionary trade-offs across environments to be explored in the laboratory[70].

Molecular springboards

Finally, we suggest three molecular mechanisms that potentiate the evolution of complex adaptations (Fig. 4). The common feature of these mechanisms is that they do not confer new molecular function. Rather, they provide access to many new evolutionary paths or reduce the number of mutations required for phenotypic change.

Gatekeepers of protein stability. A central problem in protein evolution is that mutations that cause functional change simultaneously compromise protein stability, solubility or folding rate. Therefore, the evolution of new molecular functions frequently demands prior fixation of mutations that elevate robustness to such perturbations[11]. Such permissive mutations can act within a gene or across many genes. Permissive mutations within the same protein alter global biophysical properties of the protein, and thereby mitigate the pleiotropic side-effects of function-altering mutations (Fig. 4a). They have been observed in the laboratory[35,71] as well as in nature[72]. Chaperones are often implicated in buffering harmful structure-altering mutations, and thereby facilitate accumulation of neutral mutations and molecular adaptation across the genome[73,74]. Indeed, overexpression of molecular chaperones mitigates the constraints on protein stability, and thereby improves the specificity and activity of the evolved enzymes[75]. In principle, both forms of stability-enhancing mechanisms allow the accumulation of a variety of function-changing mutations and thereby have the potential to open new evolutionary paths.

Promiscuous molecular interactions. Enzymatic side-reactions, spurious transcription factor–DNA interactions and promiscuous protein–protein interactions arise as inevitable by-products of infidelity in molecular recognition. They seem to be prevalent in metabolic, protein–protein interactions and transcriptional networks alike[76–78]. While promiscuous interactions are often weak and physiologically irrelevant, their specificity and strength can be enhanced by further mutations[79,80]. Given that they exhibit sufficient initial activities, such fortuitous interactions may provide a selective environment for the further emergence of entirely new molecular interactions and hence facilitate the establishment of multi-step pathways via adaptive walks (Fig. 4b). Recent evolution of a new metabolic pathway for the degradation of a toxic xenobiotic provides indirect evidence for this process[81]. In addition, promiscuous interactions that initially arose as by-products of evolution are also known to have contributed to the stepwise evolution of specific hormone–receptor interactions[82] and have been implicated in the

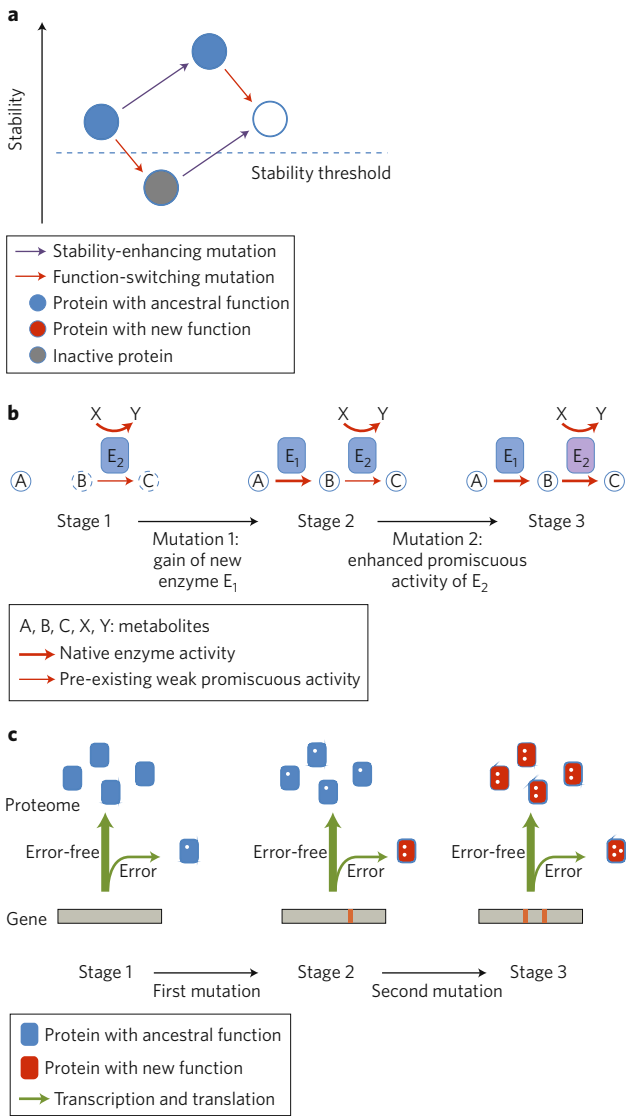

Figure 4 | Molecular springboards of complex adaptations Molecular mechanisms that potentiate the establishment of complex adaptations by eliminating fitness valleys (A) or opening up more direct mutational paths (B–C). a, Permissive mutations that increase protein stability allow the fixation of function-altering mutations that would otherwise inactivate the protein. Note that the stability-enhancing mutation might not have any fitness effect. **b**, The presence of low-level enzyme side activities facilitates the adaptive evolution of multi-step metabolic pathways. A two-step pathway with metabolites A–C is depicted where the second metabolic step can be weakly catalysed by the promiscuous side activity of enzyme E_2 (stage 1). Note that E_2 has a primary enzymatic activity outside the pathway of interest. Gain of an enzyme (E_1) catalysing the first reaction immediately confers a fitness advantage as it allows the operation of the pathway, albeit with low activity (stage 2). A second mutation enhances the side activity of E_2 and thereby results in a fully functional pathway (stage 3). **c**, Phenotypic mutations allow selection for intermediate mutations that would otherwise be neutral. The figure depicts a situation where two mutations are required for a new protein function. Owing to transcriptional and translational errors, a small fraction of the proteome already possesses one of the mutations in a non-heritable form (stage 1). A genotype carrying the other mutation thus has a selective advantage as some of its proteins will carry out the new function (stage 2). A later adaptive genetic mutation provides the full fitness benefit by converting all protein molecules within the cell from the ancestral into the new function (stage 3). Mutations/errors are depicted by white dots.

formation of new regulatory binding sites[83]. A particularly interesting corollary of the presence of promiscuous interactions concerns the co-evolution of residues participating in protein–protein interactions[80]. It is generally assumed that a disrupting mutation in one protein drives the selection of a compensatory mutation in its partner during evolution. Alternatively, interacting proteins can coevolve through the generation of promiscuous variants, which serve as mutational intermediates that preserve the ability of the two proteins for functional interaction[80].

Phenotypic heterogeneity. Infidelity of biochemical processes pervades many levels, from single molecules to pathways and cells. It has long been suggested that the resulting phenotypic heterogeneity can be a source of evolutionary adaptations[84]. A specific model concerns transcriptional and translational infidelity[85], and claims that errors speed up the evolution of complex adaptation by allowing selection for intermediate mutations that would otherwise be neutral (Fig. 4c). In a nutshell, the model proposes that when two mutations are needed for a new phenotype, the second 'mutation' is first delivered by errors during transcription and translation. Therefore, the single mutant allele will be selectively advantageous, even though it does not yet encode the complete trait. Thus, phenotypic errors allow protein sequences to 'look ahead' for a more direct path to a complex trait. The theory is appealing, as translational errors are 1,000-fold more frequent than mutations, thus generating protein variants from non-mutant genes. However, it is important to keep in mind that an elevated translational error rate has a substantial fitness cost, not least because it promotes protein aggregation[86,87]. The role of mistranslation in protein evolution has recently been studied in the laboratory. The authors showed that phenotypic mutations paved the path to what later, after gene duplication, became newly compartmentalized enzymes[84]. Thus, gene duplication followed rather than initiated the divergence of this new trait.

Future directions

Complex adaptations occur at several organizational levels. They are abundant in single proteins, protein–protein binding interfaces, metabolic pathways and transcription factor–binding site interactions. In this Review, we argued that despite the apparent functional differences, the governing evolutionary mechanisms in these systems could be similar. Perhaps the most important issue for future studies is to test non-adaptive scenarios and the impact of environmental changes on the origin of complex adaptations in nature. This ambitious goal demands discrimination of predictions. Population genetic models indicate that when the population size is large, sequential fixation of non-adaptive intermediate mutations in the population becomes exceedingly unlikely, regardless of the future benefit they may confer in combination with other alleles (Box 1). Accordingly, in large populations, complex adaptations most probably evolve either through the simultaneous fixation of mutational combinations[88] or through the sequential fixation of intermediate mutations that confer an advantage in another environmental or genetic context. These two scenarios can be discriminated by the phylogenetic and molecular reconstruction of intermediate steps of evolution and by testing the phenotypic properties of these variants. Pioneering evolutionary studies embarked on this challenge and showed that permissive mutations in a viral protein had been fixed ahead of the mutation that confers a new resistance phenotype[89]. Given the large size of virus populations, this result seems to be inconsistent with the non-adaptive model, and rather suggests that the intermediate mutations themselves were driven by selection.

Are permissive mutations generally subject to neutral or adaptive evolution? As permissive mutations in proteins often alter thermal stability, solubility or folding rate[90], they could potentially confer a fitness advantage in specific environments, such as increased temperature or ionizing radiation[90]. Recent laboratory studies unequivocally demonstrated that antagonistic pleiotropy is prevalent: mutations deleterious in one environment are beneficial in another[91]. This raises the possibility that mutations with antagonistic effects, rather than neutral mutations, contribute to the evolution of complex adaptations. A case study showed that this is indeed a realistic possibility[67]. Systematic testing of this hypothesis demands mutational effects and epistasis to be explored across environments.

We expect conceptual breakthroughs through innovative application of new molecular techniques. Deep mutational scanning[92] and new genome engineering approaches[93,94] offer unprecedented resolutions to map the genotype–fitness landscapes of single proteins and multi-gene subsystems alike. Studying mutational trajectories around adaptive peaks in many environments will give insights into the frequency by which varying environments escape adaptive stasis. Moreover, realistic fitness landscapes are highly multidimensional and may therefore contain adaptive by-passes in extra dimensions[55]. Experimental mapping of high-dimensional landscapes will provide information about the prevalence and dimensionality of such indirect paths.

These considerations have practical implications. Current practices of structure-guided protein engineering focus on mutating active sites to attain new enzymatic activities[95]. However, such mutations frequently result in complete loss of function or destabilized protein structure[96]. Addition of permissive mutations distributed through the protein structure is a prerequisite for new functional specificities. Hence, function-altering mutations should be combined with structurally non-obvious allosteric mutations (see ref. [97] for details). We anticipate that to obtain a desired function, evolutionary engineering could be facilitated by temporally varying the selection regime. Finally, we note that in computer science, standard genetic algorithms have a tendency to converge quickly to a local solution, and hence frequently fail to identify more promising regions of the search space[6]. Application of dynamically changing 'environments' offers a natural strategy to maintain the diversity required to explore the adaptive surface[98].

Received: 29 November 2016; Accepted: 2 May 2017;
Published online: 21 July 2017

References

1. Smith, J. M. *et al.* Developmental constraints and evolution: a perspective from the Mountain Lake conference on development and evolution. *Q. Rev. Biol.* **60**, 265–287 (1985).
2. Darwin, C. *On the Origin of Species* (John Murray, London, 1859).
3. Fisher, R. A. *The Genetical Theory of Natural Selection: a Complete Variorum Edition* (Oxford Univ. Press, Oxford, 1930).
4. Wright, S. Surfaces of selective value revisited. *Am. Nat.* **131**, 115–123 (1988).
5. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell. Biol.* **10**, 866–876 (2009).
6. Banzhaf, W. *et al.* Guidelines: From artificial evolution to computational evolution: a research agenda. *Nat. Rev. Genet.* **7**, 729–735 (2006).
7. Kirschner, M. & Gerhart, J. Evolvability. *Proc. Natl Acad. Sci. USA* **95**, 8420–8427 (1998).
8. Parter, M., Kashtan, N. & Alon, U. Facilitated variation: how evolution learns from past environments to generalize to new environments. *PLoS Comput. Biol.* **4**, e1000206 (2008).
9. Wagner, G. P. & Altenberg, L. Perspective: complex adaptations and the evolution of evolvability. *Evolution* **50**, 967–976 (1996).
10. Givnish, T. J. & Sytsma, K. J. *Molecular Evolution and Adaptive Radiation* (Cambridge Univ. Press, Cambridge, 2000).
11. Harms, M. J. & Thornton, J. W. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nat. Rev. Genet.* **14**, 559–571 (2013).
12. Meer, M. V., Kondrashov, A. S., Artzy-Randrup, Y. & Kondrashov, F. A. Compensatory evolution in mitochondrial tRNAs navigates valleys of low fitness. *Nature* **464**, 279–282 (2010).
13. Ivankov, D. N., Finkelstein, A. V. & Kondrashov, F. A. A structural perspective of compensatory evolution. *Curr. Opin. Struct. Biol.* **26**, 104–112 (2014).
14. Lynch, M. & Hagner, K. Evolutionary meandering of intermolecular interactions along the drift barrier. *Proc. Natl Acad. Sci. USA* **112**, E30–E38 (2015).

15. Yamada, T. & Bork, P. Evolution of biomolecular networks: lessons from metabolic and protein interactions. *Nat. Rev. Mol. Cell. Biol.* **10**, 791–803 (2009).
16. Liu, R. & Ochman, H. Stepwise formation of the bacterial flagellar system. *Proc. Natl Acad. Sci. USA* **104**, 7116–7121 (2007).
17. Pallen, M. J. & Matzke, N. J. From *The Origin of Species* to the origin of bacterial flagella. *Nat. Rev. Microbiol.* **4**, 784–790 (2006).
18. Pereira-Leal, J. B., Levy, E. D., Kamp, C. & Teichmann, S. A. Evolution of protein complexes by duplication of homomeric interactions. *Genome Biol.* **8**, R51 (2007).
19. Dietrich, M. R. From hopeful monsters to homeotic effects: Richard Goldschmidt's integration of development, evolution, and genetics. *Am. Zool.* **40**, 738–747 (2000).
20. Theiβen, G. Saltational evolution: hopeful monsters are here to stay. *Theor. Biosci.* **128**, 43–51 (2009).
21. Schrider, D., Hourmozdi, J. & Hahn, M. Pervasive multinucleotide mutational events in eukaryotes. *Curr. Biol.* **21**, 1051–1054 (2011).
22. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
23. De, S. & Babu, M. M. A time-invariant principle of genome evolution. *Proc. Natl Acad. Sci. USA* **107**, 13004–13009 (2010).
24. Hicks, W. M., Kim, M. & Haber, J. E. Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. *Science* **329**, 82–85 (2010).
25. Peisajovich, S. G., Garbarino, J. E., Wei, P. & Lim, W. A. Rapid diversification of cell signaling phenotypes by modular domain recombination. *Science* **328**, 368–372 (2010).
26. Zhang, Y. X. *et al.* Genome shuffling leads to rapid phenotypic improvement in bacteria. *Nature* **415**, 644–646 (2002).
27. Gogarten, J. P. & Townsend, J. P. Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* **3**, 679–687 (2005).
28. Pál, C., Papp, B. & Lercher, M. J. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat. Genet.* **37**, 1372–1375 (2005).
29. Sheltzer, J. M. & Amon, A. The aneuploidy paradox: costs and benefits of an incorrect karyotype. *Trends Genet.* **27**, 446–453 (2011).
30. Rancati, G. *et al.* Aneuploidy underlies rapid adaptive evolution of yeast cells deprived of a conserved cytokinesis motor. *Cell* **135**, 879–893 (2008).
31. Lynch, M. The evolution of genetic networks by non-adaptive processes. *Nat. Rev. Genet.* **8**, 803–813 (2007).
32. Amitai, G., Gupta, R. D. & Tawfik, D. S. Latent evolutionary potentials under the neutral mutational drift of an enzyme. *HFSP J.* **1**, 67–78 (2007).
33. Wagner, A. *The Origins of Evolutionary Innovations: a Theory of Transformative Change in Living Systems* (Oxford Univ. Press, Oxford, 2011).
34. Hayden, E. J., Ferrada, E. & Wagner, A. Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature* **474**, 92–95 (2011).
35. Bershtein, S., Goldin, K. & Tawfik, D. S. Intense neutral drifts yield robust and evolvable consensus proteins. *J. Mol. Biol.* **379**, 1029–1044 (2008).
36. Bloom, J. D., Romero, P. A., Lu, Z. & Arnold, F. H. Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. *Biol. Direct* **2**, 17 (2007).
37. Gray, M. W., Lukes, J., Archibald, J. M., Keeling, P. J. & Doolittle, W. F. Cell biology. Irremediable complexity? *Science* **330**, 920–921 (2010).
38. Lynch, M. The evolution of multimeric protein assemblages. *Mol. Biol. Evol.* **29**, 1353–1366 (2012).
39. Anderson, D. W., McKeown, A. N. & Thornton, J. W. Intermolecular epistasis shaped the function and evolution of an ancient transcription factor and its DNA binding sites. *eLife* **4**, e07864 (2015).
40. Finnigan, G. C., Hanson-Smith, V., Stevens, T. H. & Thornton, J. W. Evolution of increased complexity in a molecular machine. *Nature* **481**, 360–364 (2012).
41. Petrie, K. L. & Joyce, G. F. Limits of neutral drift: lessons from the *in vitro* evolution of two ribozymes. *J. Mol. Evol.* **79**, 75–90 (2014).
42. Smith, W. S., Hale, J. R. & Neylon, C. Applying neutral drift to the directed molecular evolution of a β-glucuronidase into a β-galactosidase: two different evolutionary pathways lead to the same variant. *BMC Res. Notes* **4**, 138 (2011).
43. Szappanos, B. *et al.* Adaptive evolution of complex innovations through stepwise metabolic niche expansion. *Nat. Commun.* **7**, 11607 (2016).
44. Mira, A., Ochman, H. & Moran, N. A. Deletional bias and the evolution of bacterial genomes. *Trends Genet.* **17**, 589–596 (2001).
45. Kuo, C. H. & Ochman, H. Deletional bias across the three domains of life. *Genome Biol. Evol.* **1**, 145–152 (2009).
46. Meléndez-Hevia, E., Waddell, T. G. & Cascante, M. The puzzle of the Krebs citric acid cycle: assembling the pieces of chemically feasible reactions, and opportunism in the design of metabolic pathways during evolution. *J. Mol. Evol.* **43**, 293–303 (1996).
47. Huynen, M. A., Dandekar, T. & Bork, P. Variation and evolution of the citric-acid cycle: a genomic perspective. *Trends Microbiol.* **7**, 281–291 (1999).
48. Lenski, R. E., Ofria, C., Pennock, R. T. & Adami, C. The evolutionary origin of complex features. *Nature* **423**, 139–144 (2003).
49. Notebaart, R. A., Kensche, P. R., Huynen, M. A. & Dutilh, B. E. Asymmetric relationships between proteins shape genome evolution. *Genome Biol.* **10**, R19 (2009).
50. Madan Babu, M., Teichmann, S. A. & Aravind, L. Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *J. Mol. Biol.* **358**, 614–633 (2006).
51. Costanzo, M. *et al.* A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420 (2016).
52. Chen, H. D., Jewett, M. W. & Groisman, E. A. Ancestral genes can control the ability of horizontally acquired loci to confer new traits. *PLoS Genet.* **7**, e1002184 (2011).
53. Press, M. O., Queitsch, C. & Borenstein, E. Evolutionary assembly patterns of prokaryotic genomes. *Genome Res.* **26**, 826–833 (2016).
54. Weinreich, D. M., Delaney, N. F., Depristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
55. Wu, N. C., Dai, L., Olson, C. A., Lloyd-Smith, J. O. & Sun, R. Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife* **5**, e16965 (2016).
56. Palmer, A. C. *et al.* Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. *Nat. Commun.* **6**, 7385 (2015).
57. Weinreich, D. M., Lan, Y., Wylie, C. S. & Heckendorn, R. B. Should evolutionary geneticists worry about higher-order epistasis? *Curr. Opin. Genet. Dev* **23**, 700–707 (2013).
58. Mustonen, V. & Lässig, M. From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. *Trends Genet.* **25**, 111–119 (2009).
59. Wagner, A. The white-knight hypothesis, or does the environment limit innovations? *Trends Ecol. Evol.* **32**, 131–140 (2017).
60. Kashtan, N., Noor, E. & Alon, U. Varying environments can speed up evolution. *Proc. Natl Acad. Sci. USA* **104**, 13711–13716 (2007).
61. Horowitz, N. H. On the evolution of biochemical syntheses. *Proc. Natl Acad. Sci. USA* **31**, 153–157 (1945).
62. Bell, G. Fluctuating selection: the perpetual renewal of adaptation in variable environments. *Phil. Trans. R. Soc. B* **365**, 87–97 (2010).
63. Hayden, E. J. & Wagner, A. Environmental change exposes beneficial epistatic interactions in a catalytic RNA. *Proc. R. Soc. B* **279**, 3418–3425 (2012).
64. Taute, K. M., Gude, S., Nghe, P. & Tans, S. J. Evolutionary constraints in variable environments, from proteins to networks. *Trends Genet.* **30**, 192–198 (2014).
65. Bershtein, S. & Tawfik, D. S. Ohno's model revisited: measuring the frequency of potentially adaptive mutations under various mutational drifts. *Mol. Biol. Evol.* **25**, 2311–2318 (2008).
66. Stiffler, M. A., Hekstra, D. R. & Ranganathan, R. Evolvability as a function of purifying selection in TEM-1 β-lactamase. *Cell* **160**, 882–892 (2015).
67. Steinberg, B. & Ostermeier, M. Environmental changes bridge evolutionary valleys. *Sci. Adv.* **2**, e1500921–e1500921 (2016).
68. de Vos, M. G., Dawid, A., Sunderlikova, V. & Tans, S. J. Breaking evolutionary constraint with a tradeoff ratchet. *Proc. Natl Acad. Sci. USA* **112**, 14906–14911 (2015).
69. Blount, Z. D., Borland, C. Z. & Lenski, R. E. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **105**, 7899–7906 (2008).
70. Lázár, V. *et al.* Bacterial evolution of antibiotic hypersensitivity. *Mol. Syst. Biol.* **9**, 700 (2013).
71. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proc. Natl Acad. Sci. USA* **103**, 5869–5874 (2006).
72. Gong, L. I., Suchard, M. A. & Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* **2**, e00631 (2013).
73. Bogumil, D. & Dagan, T. Cumulative impact of chaperone-mediated folding on genome evolution. *Biochemistry* **51**, 9941–9953 (2012).
74. Fares, M. A., Moya, A. & Barrio, E. GroEL and the maintenance of bacterial endosymbiosis. *Trends Genet.* **20**, 413–416 (2004).
75. Tokuriki, N. & Tawfik, D. S. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **459**, 668–673 (2009).
76. Copley, S. D. An evolutionary biochemist's perspective on promiscuity. *Trends Biochem. Sci.* **40**, 72–78 (2015).
77. Notebaart, R. A. *et al.* Network-level architecture and the evolutionary potential of underground metabolism. *Proc. Natl Acad. Sci. USA* **111**, 11762–11767 (2014).
78. Levy, E. D., Landry, C. R. & Michnick, S. W. How perfect can protein interactomes be? *Sci. Signal.* **2**, pe11 (2009).
79. Aharoni, A. *et al.* The 'evolvability' of promiscuous protein functions. *Nat. Genet.* **37**, 73–76 (2004).
80. Aakre, C. D. *et al.* Evolving new protein-protein interaction specificity through promiscuous intermediates. *Cell* **163**, 594–606 (2015).

81. Copley, S. D. Evolution of a metabolic pathway for degradation of a toxic xenobiotic: the patchwork approach. *Trends Biochem. Sci.* **25**, 261–265 (2000).
82. Bridgham, J. T., Carroll, S. M. & Thornton, J. W. Evolution of hormone-receptor complexity by molecular exploitation. *Science* **312**, 97–101 (2006).
83. Nourmohammad, A. & Lässig, M. Formation of regulatory modules by local sequence duplication. *PLoS Comput. Biol.* **7**, e1002167 (2011).
84. Yanagida, H. *et al.* The evolutionary potential of phenotypic mutations. *PLoS Genet.* **11**, e1005445 (2015).
85. Whitehead, D. J., Wilke, C. O., Vernazobres, D. & Bornberg-Bauer, E. The look-ahead effect of phenotypic mutations. *Biol. Direct* **3**, 18 (2008).
86. Kalapis, D. *et al.* Evolution of robustness to protein mistranslation by accelerated protein turnover. *PLoS Biol.* **13**, e1002291 (2015).
87. Pouplana, L. R., Santos, M. A., Zhu, J. H., Farabaugh, P. J. & Javid, B. Protein mistranslation: friend or foe? *Trends Biochem. Sci.* **39**, 355–362 (2014).
88. Weinreich, D. M. & Chao, L. Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. *Evolution* **59**, 1175–1182 (2005).
89. Bloom, J. D., Gong, L. I. & Baltimore, D. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science* **328**, 1272–1275 (2010).
90. Trudeau, D. L., Kaltenbach, M. & Tawfik, D. S. On the potential origins of the high stability of reconstructed ancestral proteins. *Mol. Biol. Evol.* **33**, 2633–2641 (2016).
91. Qian, W., Ma, D., Xiao, C., Wang, Z. & Zhang, J. The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. *Cell Rep.* **2**, 1399–1410 (2012).
92. Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods* **11**, 801–807 (2014).
93. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
94. Nyerges, Á. *et al.* A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proc. Natl Acad. Sci. USA* **113**, 2502–2507 (2016).
95. Lutz, S. Beyond directed evolution—semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* **21**, 734–743 (2010).
96. Tokuriki, N. & Tawfik, D. S. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **19**, 596–604 (2009).
97. Goldsmith, M. & Tawfik, D. S. Enzyme engineering by targeted libraries. *Methods Enzymol.* **523**, 257–283 (2013).
98. O'Neill, M., Vanneschi, L., Gustafson, S. & Banzhaf, W. Open issues in genetic programming. *Genet. Program. Evol. M* **11**, 339–363 (2010).
99. Coyne, J. A., Barton, N. H. & Turelli, M. Perspective: a critique of Sewall Wright's shifting balance theory of evolution. *Evolution* **51**, 643–671 (1997).
100. Orr, H. A. The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* **6**, 119–127 (2005).
101. Doniger, S. W. *et al.* A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet.* **4**, e1000183 (2008).
102. Henn, B. M., Botigué, L. R., Bustamante, C. D., Clark, A. G. & Gravel, S. Estimating the mutation load in human genomes. *Nat. Rev. Genet.* **16**, 333–343 (2015).
103. MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).
104. Lynch, M. & Abegg, A. The rate of establishment of complex adaptations. *Mol. Biol. Evol.* **27**, 1404–1414 (2010).
105. Lynch, M. Scaling expectations for the time to establishment of complex adaptations. *Proc. Natl Acad. Sci. USA* **107**, 16577–16582 (2010).
106. Weissman, D. B., Feldman, M. W. & Fisher, D. S. The rate of fitness-valley crossing in sexual populations. *Genetics* **186**, 1389–1410 (2010).
107. Ramani, A. K. & Marcotte, E. M. Exploiting the co-evolution of interacting proteins to discover interaction specificity. *J. Mol. Biol.* **327**, 273–284 (2003).
108. Podgornaia, A. I. & Laub, M. T. Pervasive degeneracy and epistasis in a protein–protein interface. *Science* **347**, 673–677 (2015).
109. Bridgham, J. T., Ortlund, E. A. & Thornton, J. W. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* **461**, 515–519 (2009).
110. Papp, B., Notebaart, R. A. & Pál, C. Systems-biology approaches for predicting genomic evolution. *Nat. Rev. Genet.* **12**, 591–602 (2011).

Acknowledgements

We thank D. Tawfik for his valuable comments on an earlier version of this paper. This work was supported by the 'Lendület' programme of the Hungarian Academy of Sciences (B.P. and C.P.), the Wellcome Trust (C.P. and B.P.), H202-ERC-2014-CoG (C.P.), GINOP-2.3.2-15-2016-00014 (EVOMER, C.P. and B.P.) and GINOP-2.3.2-15-2016-00020 (MolMedEx TUMORDNS, C.P.).

Author contributions

All authors contributed to writing and revising the manuscript.

Competing interests

The authors declare no competing financial interests.

Additional information

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to C.P.