Received: 28 February 2022 | Revised: 11 May 2022 | Accepted: 14 May 2022

DOI: 10.1002/pro.4362

REVIEW

Mechanisms of protein evolution

Vijay Jayaraman[1] | **Saacnicteh Toledo-Patiño**[2] | **Lianet Noda-García**[3] | **Paola Laurino**[2]

[1]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel

[2]Protein Engineering and Evolution Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan

[3]Department of Plant Pathology and Microbiology, Institute of Environmental Sciences, Robert H. Smith Faculty of Agriculture, Food and Environment, Hebrew University of Jerusalem, Rehovot, Israel

Correspondence
Paola Laurino, Protein Engineering and Evolution Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan.
Email: paola.laurino@oist.jp

Lianet Noda-García, Department of Plant Pathology and Microbiology, Institute of Environmental Sciences, Robert H. Smith Faculty of Agriculture, Food and Environment, Hebrew University of Jerusalem, Rehovot, Israel.
Email: lianet.noda@mail.huji.ac.il

Funding information
Okinawa Institute of Science and Technology Graduate University; Feinberg Graduate School, Weizmann Institute of Science; Hebrew University of Jerusalem; Okinawa Institute of Science and Technology

Review Editor: John Kuriyan

Abstract

How do proteins evolve? How do changes in sequence mediate changes in protein structure, and in turn in function? This question has multiple angles, ranging from biochemistry and biophysics to evolutionary biology. This review provides a brief integrated view of some key mechanistic aspects of protein evolution. First, we explain how protein evolution is primarily driven by randomly acquired genetic mutations and selection for function, and how these mutations can even give rise to completely new folds. Then, we also comment on how phenotypic protein variability, including promiscuity, transcriptional and translational errors, may also accelerate this process, possibly via "plasticity-first" mechanisms. Finally, we highlight open questions in the field of protein evolution, with respect to the emergence of more sophisticated protein systems such as protein complexes, pathways, and the emergence of pre-LUCA enzymes.

1 | INTRODUCTION

The first version of this manuscript was written by Paola Laurino, Lianet Noda-García, and Prof. Dan S. Tawfik, whom the authors deeply miss.

Vijay Jayaraman, Saacnicteh Toledo-Patiño contributed equally to this study.

Protein evolution encompasses a large variety of phenomena addressed by multiple disciplines including biophysics, biochemistry, and evolutionary biology. The mechanistic aspects of protein evolution may be broadly phrased as: how do changes in protein sequence occur and how do they mediate changes in protein structure, and in turn in function? Each discipline has its own angle with respect to these questions. Here, we present an

 © 2022 The Protein Society.

BOX 1 Concepts and mechanisms in protein evolution—a very brief guide

The text focuses on a few less explored aspects of protein evolution, while more established aspects are covered in this box that lists key concepts and guiding references (reviews and recent papers describing specific case studies). Scientific concepts and mechanisms are inevitably schematic (if not dogmatic). Alternative scenarios or mechanisms are denoted here side-by-side in blue (noted as "versus," "alternatively," etc.). In reality, these are not mutually exclusive and may be even complementary. Many concepts are also interrelated as indicated in our cross-referencing.

1. Transitions in protein evolution can be categorized to:

Microtransitions – Divergence of new functions while maintaining the original architecture (fold) and key active-site features (divergence within protein families and superfamilies).

Macrotransitions – Transitions between different folds including the emergence of the earliest protein folds.

2. Protein sequences diverge with time (this is what evolution means). Schematically, these changes may relate to drift or adaptation:

Drift – Sequence changes occurring due to random sampling while preserving the protein's structure and function (see *purifying selection*).

Adaptation – Changes in protein properties including the acquisition of new biochemical activities (see *positive selection*).

Selection may drive a reduction in the frequency of certain mutations (alleles) within a given population (purging, *purifying selection*) and/or the enrichment of other mutations (*positive selection*). Selection shapes protein traits including their biochemical activity (binding, catalysis, etc.) and biophysical properties (folding, stability, etc.). Traits such as enzyme selectivity relate to *positive selection*, that is, not only by enrichment of mutations that increase binding or catalytic efficiency with the target ligand/substrate but also by mutations that reduce activity with undesirable, non-cognate substrates[19,181] (see also *trade-offs*). The latter is often addressed as "*negative selection*" (although in population genetics this term is used in relation to purifying selection).

3. ***Gene duplication*** provides the raw material for new proteins. Several different mechanisms may underline the emergence of new genes via duplication.[4,31,182] Briefly, duplicated genes may evolve toward a novel function that had not been present in the ancestral, pre-duplicated gene (*neo-functionalization*). Alternatively, a bifunctional ancestor (*generalist*) may split to two *specialist* genes (*sub-functionalization*, or *divergence before duplication*). Duplication may also provide an adaptive advantage per se, by increasing protein dose and thereby augmenting a weak, pre-existing promiscuous function.[110]

4. ***Promiscuity*** relates to the coincidental pre-existence of functions that may serve as the starting point for new functions.[9–11] If such *latent*, *promiscuous* functions come under selection, they give rise to bi-functional, *generalist* intermediates. Upon gene duplication, generalist intermediates split, giving rise to two *specialists*, each performing one function (*sub-functionalization*).[14,18,21] Although duplication and going from *generalists to specialists* is a general trend, the opposing process of gene loss and/or *specialist to generalist* also occurs.[183]

5. ***Epistasis*** – The effects of mutations not only in different genes, but also within the same gene/protein can be non-additive, *that is*, *epistatic*. Epistasis has a profound impact on evolution in general, and protein evolution in particular.[184–186]

6. **Enabling/compensatory mutations** – The dominance of epistasis also means that many (probably most) mutations that eventually get fixed in evolving proteins are deleterious on their own (during *drift*, and certainly during *adaptation*). Their acceptance may therefore occur via two alternative mechanisms: A deleterious mutation transiently accumulates and is later followed by a *compensatory mutation*.[47] Alternatively, mutations that accumulate initially as neutral enable deleterious mutations to fix at a later stage (*enabling*, *permissive* mutations).[45,187,188]

Enabling and compensation (and hence epistasis) can be *local* or *specific*[184] —that is, the deleterious and enabling/compensatory mutations occur in a specific pair of residues (typically, in two contacting residues, *for example*, within active-sites) or *global*, *nonspecific* —a given mutation may enable/compensate a range of different deleterious mutations (e.g., stabilizing mutations

that may compensate many different destabilizing mutations).

7. ***Neutrality, robustness*** relates to the ability of proteins to accumulate mutations with no change of structure, stability, or function. ***Evolvability*** or ***innovability*** relate to the ability of one or a few mutations to introduce a new structure and/or function.

While seemingly contradictory, these properties are actually complementary[189,190] —this is primarily because mutations may be neutral in one context (function, environment) yet beneficial in another (e.g., neutral mutations with respect to a protein's native, physiological function may augment a latent, promiscuous activity; see also original-new function *tradeoff*).

8. ***Trade-offs in protein evolution*** – Mutations almost always affect more than one protein trait (*pleiotropy*) and often in contradictory ways. Epistasis and trade-offs are the key elements shaping the trajectories of protein evolution.[36] Several types of evolutionary trade-offs are known with respect to proteins:

Original vs *New-function trade-off* – A mutation improving a new, evolving function is likely to decrease the original one. A strong trade-off enforces *neo-functionalization*, *that is*, duplication must occur to complete divergence and specialization (*escape from adaptive conflict*).[4] In many cases, this trade-off is initially weak, thus enabling divergence toward a bifunctional, *generalist* intermediate (see *sub-functionalization* above). The magnitude of original-new trade-offs tends to vary along adaptive trajectories, starting from weak trade-offs that give rise to *generalist* intermediates and shifting to strong trade-offs as selection progresses, thus yielding a new *specialist* (typically after *duplication*).[36,95]

Stability-activity trade-off – Most mutations decrease protein stability and thereby lead to misfolding, aggregation, and/or proteolysis. New-function mutations are even more so, thus making their accumulation dependent on *enabling/compensatory mutations*.[191,192]

Folding-stability trade-off – Beyond the thermodynamic and kinetic stability of the native, folded state, the folding process itself imposes severe constrains. Trade-offs between monomer folding and assembly of oligomers or between the ability of a protein to fold and the stability of its final, folded state, may underline the birth of new proteins.[58]

Rate-accuracy trade-off – A mutation that improves the catalytic efficiency of an enzyme may reduce its selectivity. Similarly, improvement in the affinity toward the cognate ligand may also increase cross-reactivity with noncognate ligands (see also *positive* versus *negative selection*).[181]

9. ***Diminishing returns*** – Evolutionary optimizations, including protein optimizations, are subject to strong diminishing returns—early mutations confer large advantages per mutation but as the new, evolving trait improves, the improvement per mutations decreases.[36,95] *Trade-offs*, *diminishing returns*, and other factors result in many proteins being suboptimal with respect to individual traits such as catalytic efficiency, selectivity, and stability.[181,193]

10. ***Phenotypic variation*** – Variation that exists in a genetically identical population due to the noise associated with various biological processes like transcription, translation, splicing and so forth.

integrated view, through the eyes of protein scientists. We attempted to portray how multi-faceted the research of protein evolution is and discuss relatively unexplored aspects and fundamental questions that remain unanswered. However, breadth inevitably trades off with depth. Thus, we apologise if significant achievements of specific fields are not thoroughly cited.

A fundamental paradox in protein evolution is that: "nothing evolves unless it already exists," or in other words as stated by DeVries: "Natural selection may explain the survival of the fittest, but it cannot explain the arrival of the fittest."[1] Mutations, insertions/deletions, and recombination mostly induce minor changes in protein structure (*micro-transitions*) that are sufficient for the rise of new functions, although in rare cases, these can generate completely new protein folds (*macro-transitions*) (Boxes 1,*1*). Our review revolves around this classic, "Darwinian model," and covers cases where the pre-existing sequence diversity in a population give rise to new functions.

Further, we describe various mechanisms that may expedite this process. For instance, it is possible that the genomic mutations needed for conferring a novel function might not be present in a population, they can however, rise by non-genetic mechanisms mediated by errors in replication, transcription, and translation (phenotypic mutations).[2–4] Thus, the upcoming new function is

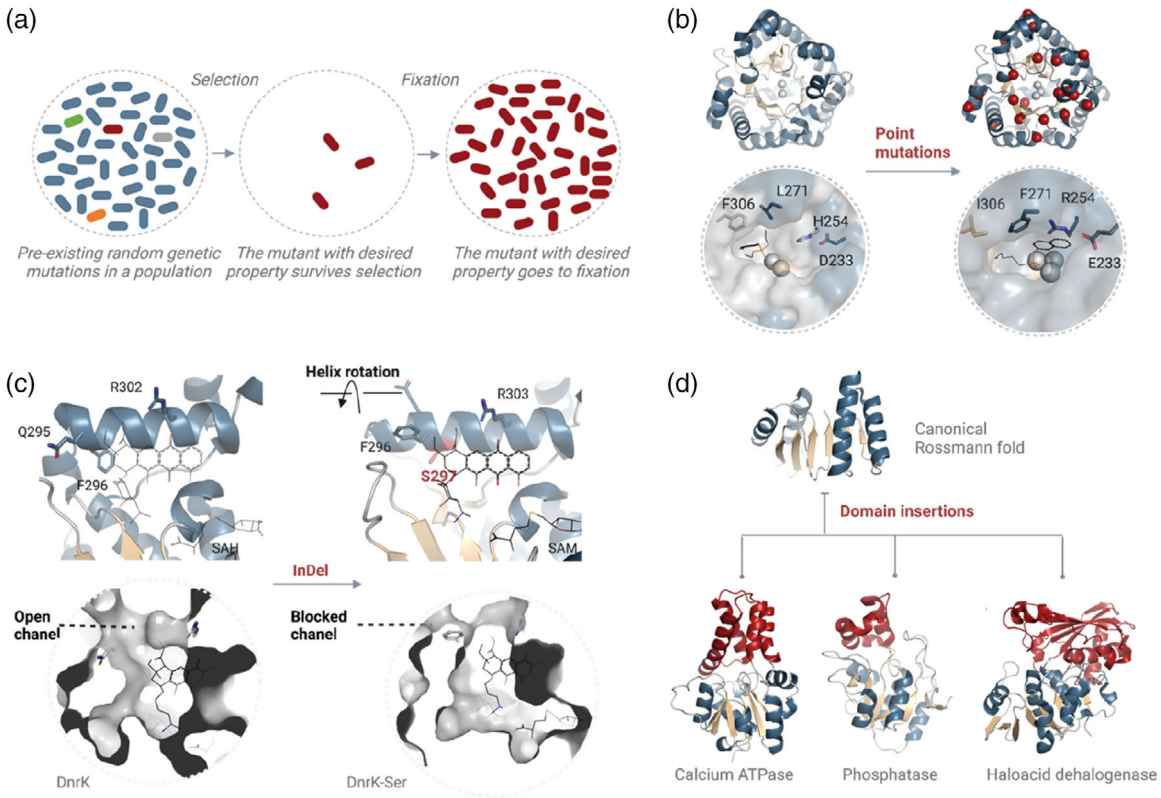

FIGURE 1 Darwinian evolution driven by pre-existing genetic changes, ranging from single amino acid mutations to gene rearrangements. (a) Schematic representation of Darwinian selection: selection purges most of the variations in the population, leading to survival of the fittest mutant, eventually undergoing fixation. (b) The outcome of a laboratory evolutionary trajectory of 18 consecutive point mutations (PDB codes: 1DPM, 2R1N, 4E3T).[36] The original and evolved active sites are depicted with their corresponding reaction intermediates (a phosphotriesterase [left] and aryl-esterase [right]). The mutated positions are denoted in red. The overall structure (cartoon) and the key catalytic residues remained unchanged (the catalytic metals are presented as grey spheres). (c) A switch between two fundamentally different activities, methyltransferase (left) and monooxygenase (right) may be triggered by an insertion of a single amino acid. An inserted serine at position 297 (red) induces a flip of the adjacent side-chain of Phe296 (blue sticks) that reshapes the active-site (surface) and triggers the activity change (PDB codes: 4WXH and 5EEG).[37] (d) Domain insertions into an existing enzyme drive the divergence of new functions,[38,39] as exemplified here for three different enzymes that share a Rossmman-fold core domain: a Haloacid dehalogenase (PDB 1ZRN), a phosphatase (1N9K), and a calcium pump-driving ATPase (1SU4). The canonical Rossmann fold is represented by a dehydrogenase (5KKA). For other examples of microtransitions see the study done by McKeown et al.; Coyle, Flores, and Lim;Bar-Rogovsky, Hugenmatter, and Tawfik; and Coelho et al.[19–22]

already present, as fortuitous, latent variation at the phenotypic level within identical genotypes (phenotypic variability).[5,6] These changes are observed at all biological levels of organization, from single proteins to entire organisms.[7,8] Indeed, the pre-existence of protein activities as latent promiscuous functions, is by now, a well-established hypothesis understood in atomic detail.[9–11] We also highlight additional aspects of phenotypic variability that underlie the arrival of the fitter. A seemingly attractive, yet controversial hypothesis, is that phenotypic variability (and possibly also genetic changes) is directly induced by environmental challenges. These so-called "Baldwin-effects"[12,13] may apply to protein evolution, and are presented here under a general model, coined "plasticity-first."

Much of the current work revolves around the evolution of individual biochemical activities such as ligand binding (DNA, RNA, small molecules, or proteins) or enzymatic functions (for recent examples see[14–23]). However, beyond biochemical activity per se, other protein features are also shaped by evolution, such as the regulation of the protein expression, folding, stability, and oligomerization[24–27] or avoiding undesired interactions with other metabolites or proteins.[28] Further, proteins also co-evolve with other proteins and biomolecules with whom they interact, and with the cellular components responsible for protein synthesis, maintenance, and clearance.[29,30] Here, we discuss some open questions related to these aspects.

As proteins have been evolving for ca. 3.7 billion years, the mechanisms underlying the divergence of

recently evolved enzymes[11,31] may appear largely inapplicable to the emergence of the very first protein(s).[32] There are, however, some unifying themes that we describe here alongside differences and unknowns. We conclude by discussing how short and functional protein fragments may have been recruited prior to the appearance of last universal common ancestor (LUCA's) proteome to give rise to primitive metabolic systems.

2 | MICROTRANSITIONS IN PROTEIN EVOLUTION

Protein evolution is driven by mutations that can occur biasedly[33,34] or at random and with no relation to selection.[35] Deleterious mutations are purged, whereas new challenges drive the fixation of mutations that give rise to proteins with modified or new functions (Boxes 1, *2*, and Figure 1).

As exemplified in Figure 1, most, if not all, of the extant protein repertoire emerged by small structural modifications while maintaining their basic fold. Such changes, dubbed *microtransitions* (Boxes 1, *1*), have been demonstrated in the laboratory, largely via point mutations, insertions/deletions (InDels), homologous or non-homologous recombination,[23] and domain fusions.[40] While the effects of point mutations have been widely explored (e.g., Figure 1b), we know less about how other types of genetic changes lead to new proteins. InDels, for example, have high adaptive potential. For instance, a single InDel can induce functional transitions[37,41] (e.g., Figure 1c). Additionally, domains frequently mix and match (gene fusion or fission) to yield new proteins.[39,42–44] The addition of a single relatively small domain allows Rossmann fold enzymes to catalyze different reactions, for example, calcium ATPase, phosphatase, and haloacid dehalogenase (Figure 1d). InDels or larger genetic rearrangements are on average, even more deleterious than point mutations and therefore intensely purged.[45,46] Acceptance of mutations in general, and InDels or larger genetic rearrangements especially, typically demands compensation by other mutations (Boxes 1, *6*).[45,47]

In contrast to the divergence of functions in existing domains, the birth of new protein topology and architecture is driven primarily by duplication and fusion of short segments, as discussed in the following section.

3 | MACROTRANSITIONS IN PROTEIN EVOLUTION

Protein domains whose secondary structural elements adopt similar orientation in space are classified under the same architecture. If in addition, these elements display identical topological connections, they are further sorted under the same fold. Substantiated cases of homology between different folds are rare. Only recently, the development of sensitive homology prediction tools has allowed drawing evolutionary bridges between folds that were previously thought unrelated.[48–55] Despite these efforts, most of the evolutionary relationships between distant homologs remain a mystery. How did the first protein folds emerge? Did transitions between these architectures occur at any stage; and if so, how?

Studies of metamorphic proteins have provided some hints,[59–62] which demonstrate that the topology and architecture of protein domains can be altered, herein called *macrotransitions*, by introducing a few or even one single amino acid substitution[56] (Boxes 1,*1*). Such is the case of the G_A protein, a serum binding domain that is converted into G_B, an IgG-binding domain upon a L45Y substitution (Figure 2a). This type of structural transition suggests the existence of critical residues that stabilize certain tertiary interactions while abolishing others. Likewise, a single protein sequence can fold into more than one structure. These sequences have more than one energetically favored minimum (scaffold plasticity) that allows the interconversion between different structures upon changes in the environment such as pH; lipid or buffer composition.[63] De novo emergence of proteins by overprinting is another example of a *macrotransition*, where alternative frames of coding sequences from short segments of existing proteins are translated. This phenomenon can give rise to new amino acid sequences, and ultimately to new protein architectures.[64–66] For instance, by incorporating an alternative start codon within the nucleocapsid protein N, an additional reading frame is created, giving rise to open-reading frames (ORF)-9b protein, which adopts a new fold (Figure 2b).[57] This process is not to be confused with the de novo emergence of proteins from non-coding DNA (see open questions), where arbitrary transcripts occasionally overlap with randomly attained ORF and become translated.[67–70] Further architectural rearrangements can emerge through trading of structurally similar regions (segment-swaps) between two or more domains, which can be found in around 13% of the PDB structures.[71] This type of *macrotransition* can also be induced by InDels within a protein sequence[53] as exemplified by the flavodoxin-like fold, which upon insertion, duplication, and fusion gave rise to a new functionality, adopting the bi-lobular hemD-like architecture (Figure 2c). Duplication and fusion of short segments can also lead to open-ended (solenoid) structures as indicated by the internal symmetry that underlines many protein folds,,[72,73] for example, αα-hairpin repeats generate

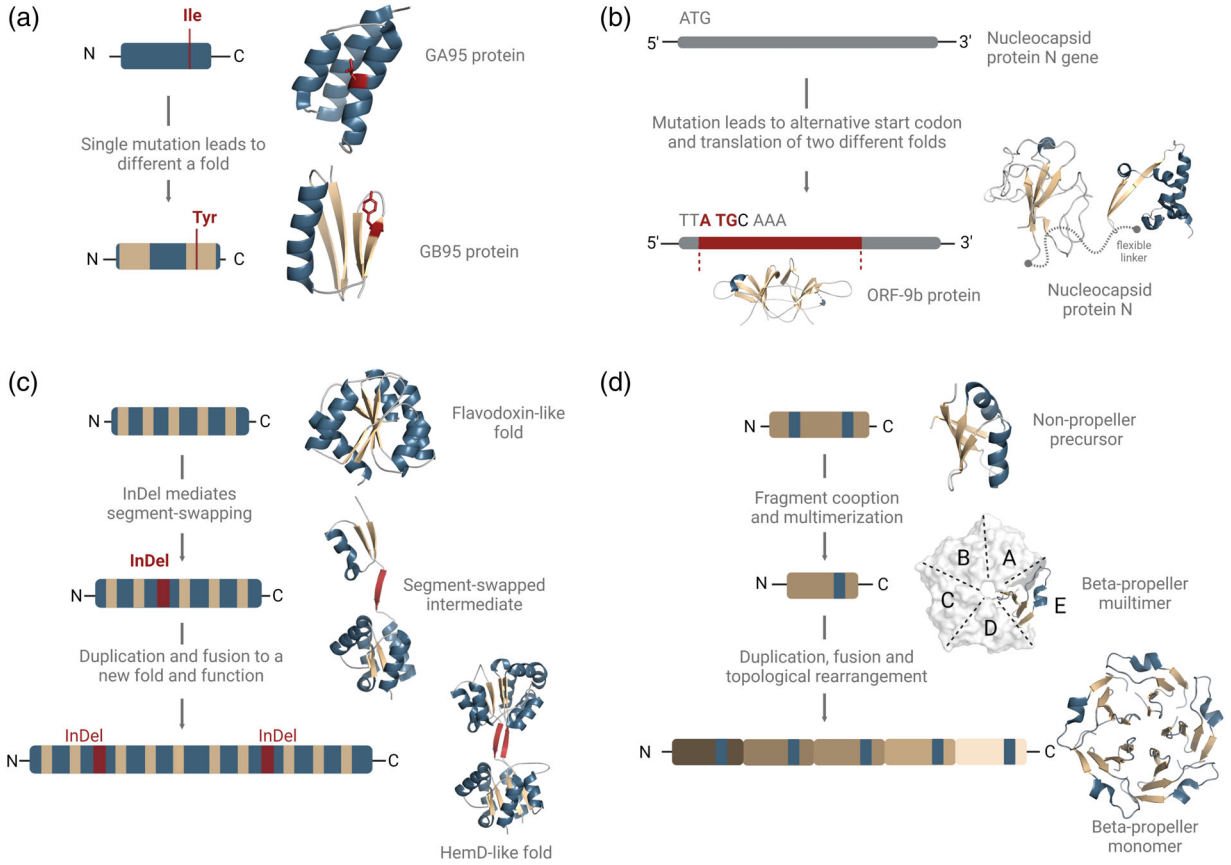

FIGURE 2 Macrotransitions: genetic mutations induce changes in protein structure. (a) A single amino acid mutation (I45Y, red) leads to a fold change as exemplified by protein GA95 (PDB 2KDL): the all-alpha structure protein acquires alpha and beta secondary elements in GB95 (PDB 2KDM). (b) Mutations at the DNA level can lead to alternative reading frames. Such is the case for the Nucleocapsid protein N gene that gives rise to the nucleocapsid N ORF-9b protein (PDB 2CME).[56] The new protein adopts an all-beta fold, in contrast to the alpha and beta elements of the original protein.[57] (c) An insertion (red) within the flavodoxin-like fold (PDB 1REQ), results in an additional beta element that segment-swaps the original fold in two. This structural rearrangement creates a protein interface that is now able to associate with another monomer, inducing the topological changes, resulting in the hemD-like fold (PDB 1jr2).[53] (d) Short fragments within proteins can act as building blocks to create novel architectures. A fragment from a non-propeller precursor (PDB 3WHI) upon oligomerization, duplication, and fusion rearrange in a monomeric propeller fold (PDB 5C2N) ORF, open-reading frames.[58]

transient receptor potential (TRP), HEAT, Armadillo, and Ankyrin structures, whereas βαβ units generate leucine-rich repeats. In other instances, repeating units create globular structures, such is the case for the triose-phosphate isomerase (TIM) barrels[74,75] and beta-propellers (Figure 2d).[58,76,77] Overall, the above-mentioned examples highlight how novel protein architectures can emerge from structurally unrelated scaffolds through relatively small changes, illustrating their plasticity and resilience potential.

While it is well known that mutations, gene rearrangements, and InDels can cause functional and structural changes in proteins, not all these mutations go to fixation. In the next section, we discuss how selection and fixation occur, based on results of various directed evolution experiments on individual proteins.

4 | SELECTION AND FIXATION OF MUTATIONS

Following their appearance, most mutations are purged while some are fixed not only by selection, but also by chance (Boxes 1, *2*). This leads to the critical question: out of all possible mutations in a protein, which fraction of these is neutral versus what fraction is deleterious and to what degree. Equally crucial is the frequency of potentially beneficial mutations and their effects on the protein's original function and stability as this dictate whether they might be fixed or rapidly purged.

The answer to these is embedded in the distribution of fitness effects (DFE) of mutations—a subject of extensive research. Systematic mappings of the effects of all possible single amino acid mutations in a given protein have become routine.[78–81] These mutational scans yield

distributions of the effects of mutations in individual proteins, and also insights regarding the structural and biochemical parameters that dictate them.[82,83] The cumulative knowledge of protein DFEs indicates that the vast majority of mutations, probably ≥80%, are deleterious,[84] with the primary reason being impaired folding and/or decreased stability.[84] Mutations that alter biochemical function are rare and also purged more intensely.[82,83] The effects of mutations on folding and stability are complex, as they also relate to how the cellular machinery deals with impaired mutants (see below). Indeed, in the short-term, mildly deleterious mutations may be tolerated owing to various cellular buffering mechanisms, thus facilitating protein evolution.[85–87]

The evolutionary interpretation of deep mutational scans is problematic, not the least because the measured "fitness" values rarely relate to organismal fitness. Accordingly, most experiments indicate higher mutational tolerance than what is observed in nature among homologous proteins, suggesting that most mutations, in laboratory conditions, do not affect structure and/or function.[84] It appears that the deleterious effects of mutations are masked in most laboratory experiments,[83] rendering the results more relevant to the understanding of short-term genetic diversity (e.g., population polymorphism), as opposed to long-term evolutionary processes.[84,88] Similarly, when it comes to adaptation (acquiring new or modified protein properties), laboratory selections may typically be too stringent, thus funnelling adaptations toward one trajectory in a limited and defined environment (a single growth medium, temperature, etc.). The gradual selection pressures and diverse environments that underlie natural evolution may shape protein adaptation in ways that differ from what has been observed in most laboratory experiments.[89,90]

5 | EVOLUTIONARY RATES OF PROTEINS

When it comes to long-term evolution, the rates by which proteins evolve vary dramatically. Even when comparing proteins of the same species or orthologues only (i.e., assuming minor changes in protein function), evolutionary rates (substitutions per site, per generation) typically span over two orders of magnitude among the proteins in the same genomes. The factors that dictate the rate of protein evolution is of major interest.[91] One key determinant is epistasis, namely, interdependency between different positions of the same gene/protein (intragenic epistasis; Boxes 1, *5*). Globular proteins in general exhibit negative epistasis (deleterious effects of two different mutations is greater than the sum of individual ones).[92] As proteins evolve, deleterious mutations can still be fixed. However, their acceptance depends on the pre-existence of other mutations (permissive, enabling mutations) or on the subsequent accumulation of compensatory mutations (Boxes 1, *6*). This context dependency of mutations dictates a slower rate of evolution. Biophysical and functional constraints also affect rates of protein evolution. These include high expression levels that make proteins more prone to aggregation and promiscuous associations and multi-functionality, thereby engage a large fraction of the protein's surface.[91,93]The latter two constraints act primarily on the protein surface, namely, surface residues that mutate fourfold faster than the core residues. Interestingly, the surface constraints slow down the divergence of other residues, in particular core residues, resulting in an overall very slow evolutionary rate.[94]

Finally, the acquisition of new functions is the strongest driving force to protein sequence changes. Accordingly, mutational trajectories that lead to new protein functions have been extensively studied, revealing in atomic detail the effects of mutations on protein structure and function (Boxes 1, *3–9*).[95] We note that nearly every long adaptive trajectory beyond few mutations, includes multiple mutations at positions distal to the active site. Despite the importance of these so-called third shell mutations their contribution to the emergence of new protein function remains poorly understood.[96,97]

6 | MUTAGENIC HOTSPOTS

Mutations that confer modified or new protein functions (adaptive mutations) may pre-exist in the population when a new challenge appears or may arise within subsequent generations —for example, both pre-existing and arising mutations have been identified in insect esterases that evolved toward insecticide resistance.[98–100] Mutations that are neutral or nearly neutral, with respect to the protein's existing function, and are therefore not purged, may become beneficial upon the emergence of a new challenge (Boxes 1, 7). Still, the occurrence of point mutations is rare (10^{-9} per site, per generation, on average). Thus, the genetic diversity available at any given moment is limited, especially in organisms with small population size. In cases where mutation(s) with adaptive potential do not pre-exist in a population, the initial response to a new challenge is critical. In this context, we review and discuss several mechanisms that may hence expedite adaptation in the absence of pre-existing genetic diversity.

Cellular stresses correlate with higher mutation rates.[101] Also, the rate and type of mutations vary dramatically, depending on local DNA context, for example, short sequence repeats[102] and in a global one for example, highly transcribed regions.[33,103] These so-called adaptive mutations arise due to high mutability of single-stranded DNA in active transcription bubbles and from replication-transcription collisions.[104,105] Similarly, highly transcribed genes may be duplicated via cDNA intermediates (retro-genes).[106] Duplications can vary from gene segments to whole genomes and may also be considered as "adaptive mutations," whenever they are stress-induced and auto-amplified.[107] Under strong selection, multiple copies of a gene mediating survival may emerge within a strikingly small number of generations and disappear immediately after selection is removed.[108,109]

Given high replication fidelity, the above-discussed mechanisms may be crucial in shortening the time gap between new challenges and the arrival of mutations that promote survival.[110] It is not trivial to establish direct causality between stress, the induction of genetic changes, and adaptation.[101] Nonetheless, their relevance is highlighted by the existence of explicitly evolved "hotspots" regions in specific genes which encode rapid heritable genetic switches, such as in surface antigen proteins of pathogenic bacteria.[102,111]

7 | PROTEIN NOISE AND PHENOTYPIC MUTATIONS

In most cases, mutations conferring new function are pre-existing in a population. Alternatively, the yet-to-become new function could be already present, as latent, coincidental phenotypic variation whereby a single genotype (a given gene sequence) may give rise to a range of protein sequences, structures, and functions, and thereby to multiple phenotypes. If phenotypic protein variability is neutral in the environment(s) under which a protein evolved—this variability comprises "molecular noise." Nonetheless, upon appearance of a new challenge, phenotypic variability may provide an immediate survival advantage and increase the adaptive potential. In proteins, phenotypic variation can be displayed in multiple ways including: (a) variable protein levels in a population of cells due to expression noise; (b) latent, promiscuous protein conformations and activities due to drift; and (c) alternative protein sequences due to transcriptional, splicing, and/or translational errors.

Here, we focus on transcriptional and translational errors. For the adaptive potential of (a) see the studies done by Rotem et al. and Garcia-Bernardo and Dunlop,[112,113] regarding (b), see Boxes 1, *4*. Translational and transcriptional errors are $\sim 10^5$ times more frequent than genetic mutations.[114,115] As a representative example, it was shown that in yeast ADH1 gene, transcriptional errors alone can affect almost every aspect of enzymes function including oligomerization, substrate binding, cofactor binding, metal binding, and post-translational modification site (Figure 3b).[116] Like genetic mutations, phenotypic mutations are not limited to single amino acid exchanges—frameshifts, alternative starts and/or stops codons, and larger rearrangements (e.g., via alternative splicing) are also common (Figure 3b-d). Overall, given the wealth of noise associated with transcription, mRNA processing and protein synthesis, protein copies that deviate from the expected translated gene sequence are abundant.[114,117,118] These so-called phenotypic mutations have a role in the evolutionary shaping of proteins,[2,119] and may also provide starting points for emergence of new proteins or functions.[115,120] Short peptide segments that result from "illegitimate" translation small ORFs (smORFs) are also prevalent due to alternative start/stop codons, off-frame translation of coding sequences, or translation of complementary strands or even of noncoding regions.[65,121–123] Such segments might also comprise the starting material for novel proteins (Figure 3c, d).[124,125]

It is important to highlight the fact that, although phenotypic mutations are not heritable as such, the potential to make them can be.[118,126–129] For example, the frequency of transcriptional/translational errors is highly variable and sequence dependent. Codon usage strongly affects the rate of mistranslation.[130] The frequency of slippage to yield phenotypic frameshifts is directly proportional to repeat length, eight consecutive A's being an example of programed slippage.[131–133] Therefore, selection may favor gene sequences that increase the frequency of alternative protein variants while retaining the original wild-type protein sequence. In this manner, errors that occur largely at random can be amplified at specific sites and can also be heritable.

Although phenotypic mutations occur at higher frequency and are shown to be important for the adaptation of organisms,[134–137] the experimental evidence for their adaptive role in protein evolution is only recently emerging. Direct evidence for the evolutionary role of phenotypic mutations came from the emergence of a new yeast enzyme paralogue (Figure 3d).[120] The ancestral gene of isocitrate dehydrogenase (IDP) encodes two enzyme forms (isozymes)—a major cytosolic form by intact translation and a minor form that possess a C-terminal peroxisomal signal peptide due to a translational frameshift. Following duplication, a single nucleotide genetic deletion gave rise to a new, legitimate peroxisomal paralogue,

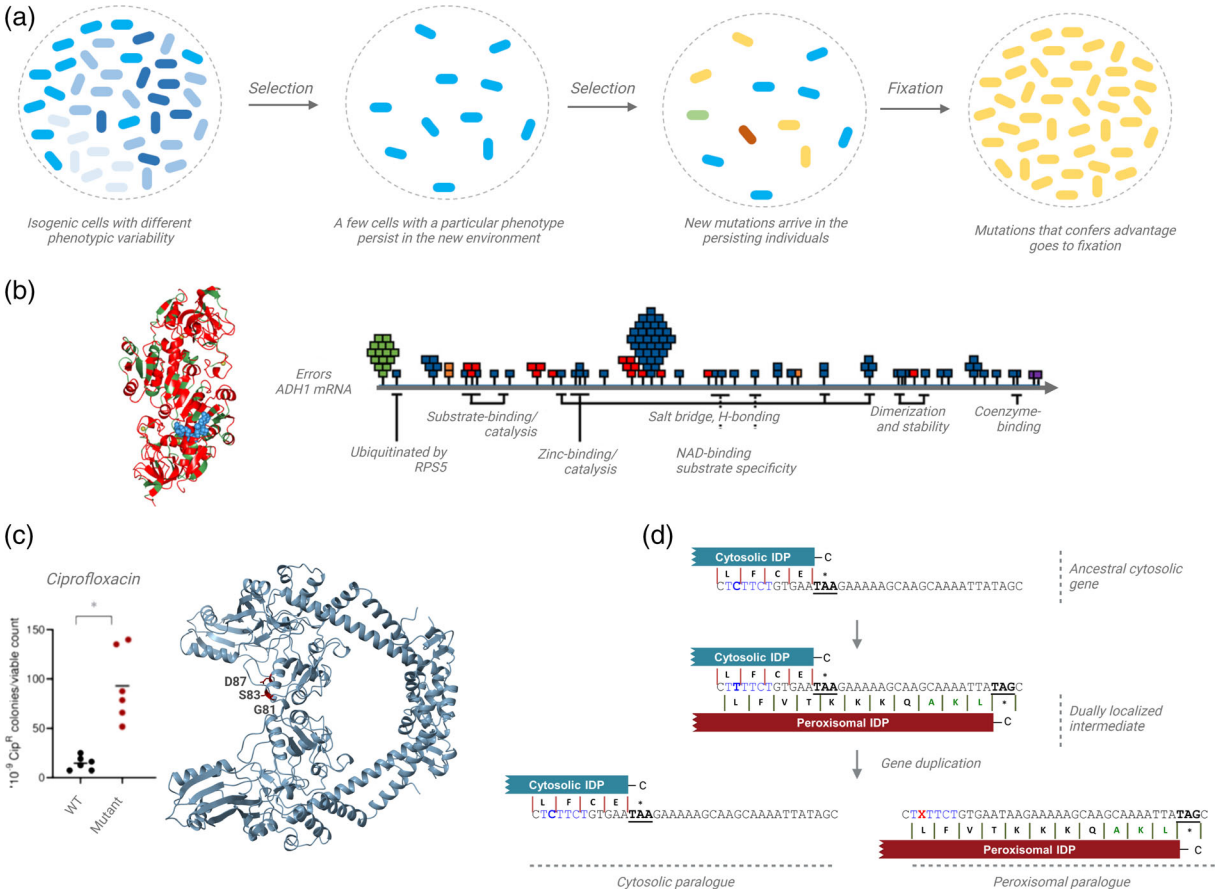

FIGURE 3 Plasticity-first mechanisms driving protein evolution (a) A schematic representation of selection that follows the plasticity-first mechanism. A new environmental challenge selects a subset of phenotypically variable isogenic cells. The phenotype permits the survival of cells, providing time for the occurrence of a mutation which confers an adaptive advantage. The mutant cells take over the entire population (fixation). (b) Transcriptional errors in yeast ADH1 mRNA mapped on to the structure. The residues with errors are highlighted in red. The scheme at the bottom shows that the mutations can affect several aspects of the enzyme: oligomerization, substrate binding, cofactor binding, metal binding, and post-translational modification sites.[116] The figure panel is reprinted from the reference (116). (c) An *Escherichia coli* mutant exhibiting higher mistranslation rates (phenotypic plasticity) displays higher frequency of genetic mutations that confer antibiotic resistance (adaptation). This panel is reprinted from the reference (135). The right panel shows the structure of DNA gyrase with the mutations conferring ciprofloxacin resistance highlighted in red. (d) Translational errors may provide the raw material to new proteins.[120] In the depicted example, a coincidental translational slippage at a TCTTTT site produces an alternative protein form with a C-terminal peroxisomal signal. In the second step, a mutation of C-to-T, that is silent with respect to the original frame, increases slippage rate, thus generating two alternative protein forms from one gene: the original cytosolic form and a minor peroxisomal form (the AKL peroxisomal signal peptide, denoted in green). Finally, following gene duplication, a single base deletion gives rise to a new, legitimate peroxisomal paralogue, whereas the original, cytosolic gene loses the cryptic peroxisomal signal.

whereas the cytosolic paralogue lost the translational frameshift that leads to a peroxisomal signal.

8 | GENETIC ACCOMMODATION OF PHENOTYPIC MUTATIONS

Phenotypic mutations may bridge the time gap between the appearance of a new challenge and the emergence of a mutation that resolves it (a gap that can be much longer than intuitively assumed). If a challenge persists, what initially comprises coincidental noise often becomes a "legitimate" function via the fixation of mutations at the genomic level that refine this function. For example, typically following gene duplication a weak promiscuous enzymatic activity may increase in both rate and selectivity to become the primary function. This was demonstrated recently in a study employing *Escherchia coli* strains with varying levels of translation error rates. The authors show that the *E. coli* mutants with higher error rates show higher frequency of ciprofloxacin resistant colonies compared to WT strains (Figure 3c).[138] Accordingly, lowering the mistranslation rates, reduced the frequency of resistant colonies as well. It is worth noting

that the genotypic mutation is often different from the phenotypic mutation.

Promiscuous protein activities seem to have a unique evolutionary advantage—mutations that increase them usually have either weak or no deleterious effects on the protein's primary activity (Boxes 1, *4*, *7*, *8*). Phenotypic mutations may also have a unique advantage in how they are genetically accommodated. In the yeast IDP case described above, single base deletions that accommodate the new trait at the DNA level (i.e., in-frame translation of the peroxisomal signal to direct all protein molecules to the peroxisome) occur at the very same mRNA site at which translational slippage occurs.[120] Overlaps between sites of genetic and phenotypic mutations have also been observed in an in vitro study.[131] Thus, selection of genotypes exhibiting a higher rate of a specific phenotypic mutation also gives rise to a hotspot for genetic mutations that accommodate the very same trait.[120,131] Similarly, ambiguous decoding (translation of a given codon to two different amino acids) was genetically accommodated in certain organisms via divergence of a dedicated tRNA.[117] More recently, it was also shown that phenotypic mutation can reduce the mutational load in a population by efficiently purging deleterious mutation. Accordingly, phenotypic mutation exhibits negative epistasis with DNA or genotypic mutation.[139]

9 | PLASTICITY-FIRST: AN EMERGING MODEL FOR PROTEIN EVOLUTION

The so-called Baldwin-effect[12] or in its more modern form, the "plasticity-first" model[13] refers to the phenomena when non-hereditary molecular variability induced by an environmental change enables initial survival. This buys time for the emergence and accommodation of genetic mutations, ensuring long-term survival of the population in the new environment. Both phenotypic plasticity and the ensuing genetic accommodation of mutations have been extensively examined and debated in the context of developmental plasticity and evolutionary adaptations.[8] Here, we adapted the Baldwin effect[12] and following a recent and insightful review,[13] present the key criteria for such a mechanism to be applied to protein evolution (Figure 3a).

The most critical criterion for proving the "plasticity-first" model for protein evolution is that the yet-to-evolve trait becomes more variable in response to the physiological stress that accompanies the new challenge. For example, the magnitude of certain promiscuous activities or the frequency of translational errors may increase due to changes in metabolite concentrations or pH. Similarly, if some pre-existing, cryptic genetic variation happens to increase the magnitude of trait variability, this would of course promote the "plasticity-first" mechanism. This criterion is not trivial to establish, and to the best of our knowledge, has not been directly examined in relation to a proven case of protein evolution.

Indeed, in many cases where the history of acquisition of new protein functions were tracked down; pre-existing promiscuous functions[140,141] or phenotypic mutations[115,120] were found to have been starting points and even provide initial survival of the population.[142,143] However, such trajectories may be perfectly accounted by a Darwinian mutation-selection model, since the pre-existence of mutants with an optimal activity in the population was not examined. Therefore, a key challenge remains to show that the latent activity was present at a sufficient level to provide a selective advantage before genetic accommodation of mutations.

Increased molecular noise is inevitably associated with reduced fitness. The cost of increased rate of translational errors, may be tolerable in short-term,[114] but in the long-term, high error rates rarely persist.[101] Overall, while the "plasticity-first" model presents an elegant shortcut to the "arrival of the fitter," direct evidence for its role in protein functional evolution is yet to be provided.

10 | PROTEIN EVOLUTION—BEYOND BIOCHEMICAL ACTIVITY

Biochemical activity—be it ligand binding or catalysis—is the primary driving force of evolutionary innovation. However, within their natural context, proteins are shaped by additional needs and forces that are complex (see Boxes 1, *8*). Following their translation, proteins fold into their native state, and must be sufficiently stable to avoid misfolding, aggregation, and/or proteolysis. The interactions of proteins with the cellular machineries that control protein quality are therefore crucial. Chaperones and also proteases, therefore impacts the type and number of mutations tolerated and thus impacts protein evolution.[85,86,144]

Regulation of protein expression is another key property shaping evolutionary trajectory. As indicated by their faster sequence divergence, non-coding elements are more evolvable than the proteins they regulate.[145] Often, the initial steps, and even the driving force for divergence may involve a new mode of transcriptional regulation. Further, the divergence of a new biochemical function is often initiated by increase in expression of an existing protein with a latent, promiscuous function.[146] This divergence may occur via mutations in the gene's own

promoter, in genes encoding other regulatory elements or via gene duplication (Boxes 1, *3*). By the current view, most new genes and paralogues especially, diverged in their transcriptional regulation and not in their biochemical function.[147,148] A classic example is the divergence by duplication of yeast Gal1/3. The ancestral, pre-duplicated gene, Gal1, encoded an enzyme, b-galactosidase that also acts as transcriptional co-inducer. Upon duplication, the new paralogue, Gal3, specialized as co-inducer, primarily via changes in the promoter that enabled faster triggering of Gal1's transcription upon appearance of lactose.[146]

Changes in the regulation of protein expression can also affect the evolvability of proteins. In fact, expression levels and protein concentration correlate with evolutionary rates—the higher the protein amount in the cell, the slower the rate,[91] although to our knowledge, direct causality has not been established. In the case of Gal3 although the key adaptive step was due to the changes in the promoter,[146] protein activity was also changed. Specialized as a co-inducer, Gal3 lost its enzymatic activity, but gain the ability to bind to Gal80 (the transcriptional repressor) with >10-fold higher affinity compared to Gal1, thus providing a distinct advantage upon switching to lactose as carbon source.[149] The divergence of new genes therefore involves changes in gene expression, that in turn enable changes in protein activity, and vice versa—in other words, noncoding and coding regions coevolve.[149,150] Beyond transcription, levels of translation are regulated, as are cellular protein levels (via changes in protein turnover rates). The mechanisms and dynamics behind the coevolution of protein expression, turnover, and function remain to be elucidated.

Proteins seldom work as independent subunits, and often self-assemble (homomers) or associated with other proteins (heteromers). About 60% of proteins are known to form complexes.[151] How these multimeric assemblies emerge and if there is adaptive value for these complexes is not clear. Recent experimental[152] and theoretical work[151] suggest that these complexes can emerge by neutral drift just like in the case of catalytic promiscuity. Often, as little as one or two mutations are enough to form new homomeric complexes.[153] Though it is tempting to associate an adaptive value for these assemblies, this remains to be investigated. Finally, protein evolution is also constrained by its cellular location. A new localization imposes new challenges. Approximately, 30% of the yeast paralogues and ~15% of *Arabidopsis* paralogues diverged in localization.[154,155] Beyond retargeting, typically by the acquisition of a signal peptide,[120] a change in localization enforces adaptation toward export (that may involve unfolding and refolding), different pH and/or redox state, and new protein partners. Overall, protein adaptation is a comprehensive process involving multiple parameters in addition to biochemical activity. Foremost, it is a process of coevolution involving the protein itself, its transcriptional and translational regulatory elements, the cellular protein-handling machineries, and other proteins and biomolecules that interact with the evolving protein.

11 | OPEN QUESTIONS

Beyond the series of questions mentioned above, there are, in our view, three key aspects in protein evolution that remain largely unanswered.

Multiple, interlocked protein components. Proteins rarely confer physiological advantage on their own. Typically, they are part of a system—a pathway or whole network involving several proteins—whereby loss of any one of these proteins results in loss of function of the entire system. For example, biosynthetic pathways comprise several enzymes, and loss of any of which of these enzymes typically results in no product. How did these multiple interlocked protein systems (MIPSs) emerge in the first instance?

Many MIPSs can be unlocked—suffice to say that free-living natural bacteria with <1,400 genes are known, and even these genomes can probably be reduced.[156] Thus, the current state of a MIPS does not reflect its initial, emergent state. Relatively simple scenarios for the emergence of MIPSs have been hypothesized.[32] With respect to metabolic pathways, bifunctional enzymes are commonly found, suggesting that certain pathways may have a priori evolved to catabolize more than one nutrient, or produce more than one product, and at later stages diverged and specialized (Box 1).[157–159] Nonetheless, the emergence of the first MIPSs, and specifically of the core biosynthetic pathways, remains enigmatic. Spontaneous occurrence of reactions, alongside a few multifunctional enzymes, may have enabled the formation of key metabolites, thus seeding the future pathways.[160–163]

Pre-LUCA recruitment of the first enzymes. In the pre-LUCA world, modern enzymes did not exist. Rather, ribozymes, metals, and H^+ and OH^- ions[164] may have been the principal catalysts. In this scenario, it has been postulated that the first peptides could have emerged to assist these early catalysts.[165,166] In fact, the exceptional abilities of peptides to chelate metals, catalyze reaction by themselves, and concentrate in condensate to enhance their activity, make them ideal seeds for the emergence of complex enzymes.[167,168] An alternative scenario includes amyloids as plausible catalytic unit at the origin of life.[170,171] Not only they show an extraordinary stability against UV radiation, different pH, and high salt concentrations, but they also catalyze diverse reactions,

including their own formation and correction while being replicated. For these reasons, the catalytic role of prototype peptides and/or amyloids prior to the putative pre-LUCA world cannot be excluded.[169] An early form of metabolism could have started via the recruitment of small peptides with catalytic properties. These units can be seen as minimalistic representations of enzymes.[172–174] Sequence and structural studies on protein domains suggest that the first proteins may have emerged by repetition, fusion, recombination, and augmentation of primordial peptides.[175] These peptide units can be found in modern protein domains with distinct global architecture[55,176,177] and were probably catalytically active as stand-alone, even if less efficient than their contemporary descendants, as well as stable enough to survive. Many questions remain unanswered on how these minimal and functional structures were recruited to replace pre-biotic catalysts and eventually lead to modern protein world.

De novo *emergence of proteins*. So far, we have addressed a large body of evidence related to transitions (micro- and macro-) in proteins that have a pre-existing globular 3D-structure (and function), but how does structure and function evolve in de novo proteins? De novo proteins are encoded in genes that emerge from non-coding segments of the DNA sequence.[178–180] These new proteins are highly disordered and represent an excellent model system to study how globular proteins evolved from a disordered precursor. The foldability of a de novo protein was examined in detail, showing that it adopts a rudimentary fold, exhibits amyloid-like propertiesm, and could act as a precursor for the emergence of fully folded proteins.[179] The study of de novo proteins might provide in the future new general principles for the evolution of folded proteins.

Overall, the evolution of MIPSs, the recruitment of first enzymes, and de novo emergence of proteins are aspects where our knowledge is still at infancy. As our understanding of how proteins evolve advances, new insights will emerge that address these and other key questions.

AUTHOR CONTRIBUTIONS

Vijay Jayaraman: Writing – review and editing (equal). **Saacnicteh Toledo-Patino:** Writing – review and editing (equal). **Lianet Noda-Garcia**: Writing-review and editing (equal). **Paola Laurino:** Writing – review and editing (equal).

ACKNOWLEDGMENTS

We thank all members of Dan S Tawfik's lab for insightful discussions on the first version of this manuscript. We thank Vikram Alva, Yitzhak Pilpel, Amy Stanton Gooch, and Raul Mireles for the critical reading of the manuscript. Financial support from the Okinawa Institute of Science and Technology to Paola Laurino and from the Hebrew University of Jerusalem to Lianet Noda-García is gratefully acknowledged. Vijay Jayaraman is a senior post-doctoral fellow from the Feinberg Graduate School, Weizmann Institute of Science.

ORCID

Vijay Jayaraman https://orcid.org/0000-0001-6288-1841
Saacnicteh Toledo-Patiño https://orcid.org/0000-0003-0632-0075
Lianet Noda-García https://orcid.org/0000-0002-0449-0700
Paola Laurino https://orcid.org/0000-0002-3725-2645

REFERENCES

1. de Vries H. Species and varieties: Their origin by mutation: Lectures delivered at the University of California. Chicago, USA: Open Court Publishing Company, 1904.
2. Goldsmith M, Tawfik DS. Potential role of phenotypic mutations in the evolution of protein expression and stability. Proc Natl Acad Sci U S A. 2009;106(15):6197–6202. https://doi.org/10.1073/pnas.0809506106.
3. Whitehead DJ, Wilke CO, Vernazobres D, Bornberg-Bauer E. The look-ahead effect of phenotypic mutations. Biol Direct. 2008;3:1–15. https://doi.org/10.1186/1745-6150-3-18.
4. Sikosek T, Chan HS, Bornberg-Bauer E. Escape from adaptive conflict follows from weak functional trade-offs and mutational robustness. Proc Natl Acad Sci U S A. 2012;109(37):14888–14893. https://doi.org/10.1073/pnas.1115620109.
5. Ackermann M. A functional perspective on phenotypic heterogeneity in microorganisms. Nat Rev Microbiol. 2015;13(8):497–508. https://doi.org/10.1038/nrmicro3491.
6. Wagner A. The origins of evolutionary innovations: A theory of transformative change in living systems, Oxford University Press Inc., New York, 2011. https://doi.org/10.1093/acprof:oso/9780199692590.001.0001
7. Tawfik DS. Messy biology and the origins of evolutionary innovations. Nat Chem Biol. 2010;6(10):692–696. https://doi.org/10.1038/nchembio.441.
8. Mary Jane west-Eberhard. Developmental Plasticity and Evolution; 2003. https://doi.org/10.1093/oso/9780195122343.001.0001
9. Pandya C, Farelli JD, Dunaway-Mariano D, Allen KN. Enzyme promiscuity: Engine of evolutionary innovation. J Biol Chem. 2014;289(44):30229–30236. https://doi.org/10.1074/jbc.R114.572990.
10. Copley SD. An evolutionary biochemist's perspective on promiscuity. Trends Biochem Sci. 2015;40(2):72–78. https://doi.org/10.1016/j.tibs.2014.12.004.
11. Khersonsky O, Tawfik DS. Enzyme promiscuity: A mechanistic and evolutionary perspective. Annu Rev Biochem. 2010;79:471–505. https://doi.org/10.1146/annurev-biochem-030409-143718.
12. Crispo E. The Baldwin effect and genetic assimilation: Revisiting two mechanisms of evolutionary change mediated by

phenotypic plasticity. Evolution. 2007;61(11):2469–2479. https://doi.org/10.1111/j.1558-5646.2007.00203.x.
13. Levis NA, Pfennig DW. Evaluating "plasticity-first" evolution in nature: Key criteria and empirical approaches. Trends Ecol Evol. 2016;31(7):563–574. https://doi.org/10.1016/j.tree.2016.03.012.
14. Hudson WH, Kossmann BR, de Vera IMS, et al. Distal substitutions drive divergent DNA specificity among paralogous transcription factors through subdivision of conformational space. Proc Natl Acad Sci U S A. 2016;113(2):326–331. https://doi.org/10.1073/pnas.1518960113.
15. Escudero JA, Loot C, Parissi V, Nivina A, Bouchier C, Mazel D. Unmasking the ancestral activity of integron integrases reveals a smooth evolutionary transition during functional innovation. Nat Commun. 2016;7:10937. https://doi.org/10.1038/ncomms10937.
16. Salmon M, Laurendon C, Vardakou M, et al. Emergence of terpene cyclization in Artemisia annua. Nat Commun. 2015;6:4–13. https://doi.org/10.1038/ncomms7143.
17. Elliott AG, Delay C, Liu H, et al. Evolutionary origins of a bioactive peptide buried within Preproalbumin. Plant Cell. 2014;26(3):981–995. https://doi.org/10.1105/tpc.114.123620.
18. Pougach K, Voet A, Kondrashov FA, et al. Duplication of a promiscuous transcription factor drives the emergence of a new regulatory network. Nat Commun. 2014;5:4868. https://doi.org/10.1038/ncomms5868.
19. McKeown AN, Bridgham JT, Anderson DW, Murphy MN, Ortlund EA, Thornton JW. Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. Cell. 2014;159(1):58–68. https://doi.org/10.1016/j.cell.2014.09.003.
20. Coyle SM, Flores J, Lim WA. XExploitation of latent allostery enables the evolution of new modes of MAP kinase regulation. Cell. 2013;154(4):875–887. https://doi.org/10.1016/j.cell.2013.07.019.
21. Bar-Rogovsky H, Hugenmatter A, Tawfik DS. The evolutionary origins of detoxifying enzymes: The mammalian serum paraoxonases (PONs) relate to bacterial homoserine lactonases. J Biol Chem. 2013;288(33):23914–23927. https://doi.org/10.1074/jbc.M112.427922.
22. Coelho PS, Wang ZJ, Ener ME, et al. A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo. Nat Chem Biol. 2013;9(8):485–487. https://doi.org/10.1038/nchembio.1278.
23. Trudeau DL, Smith MA, Arnold FH. Innovation by homologous recombination. Curr Opin Chem Biol. 2013;17(6):902–909. https://doi.org/10.1016/j.cbpa.2013.10.007.
24. Agozzino L, Dill KA. Protein evolution speed depends on its stability and abundance and on chaperone concentrations. Proc Natl Acad Sci U S A. 2018;115(37):9092–9097. https://doi.org/10.1073/pnas.1810194115.
25. Manhart M, Morozov AV. Protein folding and binding can emerge as evolutionary spandrels through structural coupling. Proc Natl Acad Sci USA. 2015;112(6):1797–1802. https://doi.org/10.1073/pnas.1415895112.
26. Modi T, Campitelli P, Kazan IC, Ozkan SB. Protein folding stability and binding interactions through the lens of evolution: A dynamical perspective. Curr Opin Struct Biol. 2021;66:207–215. https://doi.org/10.1016/j.sbi.2020.11.007.
27. Levy E, Teichmann S. Chapter Two - Structural, Evolutionary, and Assembly Principles of Protein Oligomerization. Prog Mol Biol Transl Sci. 2013;117:25–51. https://doi.org/10.1016/B978-0-12-386931-9.00002-7.
28. Castro-Fernandez V, Herrera-Morande A, Zamora R, et al. Reconstructed ancestral enzymes reveal that negative selection drove the evolution of substrate specificity in ADP-dependent kinases. J Biol Chem. 2017;292(38):15598–15610. https://doi.org/10.1074/jbc.M117.790865.
29. Hartl FU, Bracher A, Hayer-Hartl M. Molecular chaperones in protein folding and proteostasis. Nature. 2011;475(7356):324–332. https://doi.org/10.1038/nature10317.
30. Chen B, Retzlaff M, Roos T, Frydman J. Cellular strategies of protein quality control. Cold Spring Harb Perspect Biol. 2011;3(8):1–14. https://doi.org/10.1101/cshperspect.a004374.
31. Soskine M, Tawfik DS. Mutational effects and the evolution of new protein functions. Nat Rev Genet. 2010;11(8):572–582. https://doi.org/10.1038/nrg2808.
32. Noda-Garcia L, Liebermeister W, Tawfik DS. Metabolite-enzyme coevolution: From single enzymes to metabolic pathways and networks. Annu Rev Biochem. 2018;87:187–216. https://doi.org/10.1146/annurev-biochem-062917-012023.
33. Monroe JG, Srikant T, Carbonell-Bejerano P, et al. Mutation bias reflects natural selection in Arabidopsis thaliana. Nature. 2022;602:101–105. https://doi.org/10.1038/s41586-021-04269-6.
34. McDonald J, Kreitman M. Adaptive protein evolution at Adh in drosophila. Nature. 1991;351:652–654. http://ib.berkeley.edu/labs/slatkin/popgenjclub/pdf/mcdonald-kreitman1991.pdf.
35. Luria SE, Delbruck M. Mutations of bacteria from virus sensitivity to virus resistance. Genetics. 1943;28:491–511.
36. Kaltenbach M, Tokuriki N. Dynamics and constraints of enzyme evolution. JEZ-B Molecular and Developmental Evolution. 2014;322(7):468–487. https://doi.org/10.1002/jez.b.22562.
37. Grocholski T, Dinis P, Niiranen L, Niemi J, Metsä-Ketelä M. Divergent evolution of an atypical S-adenosyl-L-methionine-dependent monooxygenase involved in anthracycline biosynthesis. Proc Natl Acad Sci USA. 2015;112(32):9866–9871. https://doi.org/10.1073/pnas.1501765112.
38. Meng EC, Babbitt PC. Topological variation in the evolution of new reactions in functionally diverse enzyme superfamilies. Curr Opin Struct Biol. 2011;21(3):391–397. https://doi.org/10.1016/j.sbi.2011.03.007.
39. Allen KN, Dunaway-Mariano D. Markers of fitness in a successful enzyme superfamily. Curr Opin Struct Biol. 2009;19(6):658–665. https://doi.org/10.1016/j.sbi.2009.09.008.
40. Pandya C, Brown S, Pieper U, et al. Consequences of domain insertion on sequence-structure divergence in a superfold. Proc Natl Acad Sci USA. 2013;110(36):3381–3387. https://doi.org/10.1073/pnas.1305519110.
41. Tóth-Petróczy Á, Tawfik DS. Hopeful (protein InDel) monsters? Structure. 2014;22(6):803–804. https://doi.org/10.1016/j.str.2014.05.013.
42. Moore AD, Björklund ÅK, Ekman D, Bornberg-Bauer E, Elofsson A. Arrangements in the modular evolution of proteins. Trends Biochem Sci. 2008;33(9):444–451. https://doi.org/10.1016/j.tibs.2008.05.008.

43. Dohmen E, Klasberg S, Bornberg-Bauer E, Perrey S, Kemena C. The modular nature of protein evolution: Domain rearrangement rates across eukaryotic life. BMC Evol Biol. 2020;20(1):1–13. https://doi.org/10.1186/s12862-020-1591-0.
44. Weiner J, Beaussart F, Bornberg-Bauer E. Domain deletions and substitutions in the modular protein evolution. FEBS J. 2006;273(9):2037–2047. https://doi.org/10.1111/j.1742-4658.2006.05220.x.
45. Tóth-Petróczy Á, Tawfik DS. Protein insertions and deletions enabled by neutral roaming in sequence space. Mol Biol Evol. 2013;30(4):761–771. https://doi.org/10.1093/molbev/mst003.
46. Emond S, Petek M, Kay EJ, et al. Accessing unexplored regions of sequence space in directed enzyme evolution via insertion/deletion mutagenesis. Nat Commun. 2020;11(1):1–14. https://doi.org/10.1038/s41467-020-17061-3.
47. Leushkin EV, Bazykin GA, Kondrashov AS. Insertions and deletions trigger adaptive walks in drosophila proteins. Proc R Soc B Biol Sci. 2012;279(1740):3075–3082. https://doi.org/10.1098/rspb.2011.2571.
48. Grishin NV. Fold change in evolution of protein structures. J Struct Biol. 2001;134(2–3):167–185. https://doi.org/10.1006/jsbi.2001.4335.
49. Andreeva A, Murzin AG. Evolution of protein fold in the presence of functional constraints. Curr Opin Struct Biol. 2006;16(3):399–408. https://doi.org/10.1016/j.sbi.2006.04.003.
50. Andreeva A, Prlić A, Hubbard TJP, Murzin AG. SISYPHUS - structural alignments for proteins with non-trivial relationships. Nucleic Acids Res. 2007;35(1):253–259. https://doi.org/10.1093/nar/gkl746.
51. Alva V, Koretke KK, Coles M, Lupas AN. Cradle-loop barrels and the concept of metafolds in protein classification by natural descent. Curr Opin Struct Biol. 2008;18(3):358–365. https://doi.org/10.1016/j.sbi.2008.02.006.
52. Alva V, Remmert M, Biegert A, Lupas AN, Söding J. A galaxy of folds. Protein Sci. 2010;19(1):124–130. https://doi.org/10.1002/pro.297.
53. Toledo-Patiño S, Chaubey M, Coles M, Höcker B. Reconstructing the remote origins of a fold singleton from a Flavodoxin-like ancestor. Biochemistry. 2019;58(48):4790–4793. https://doi.org/10.1021/acs.biochem.9b00900.
54. josé arcadio Farías-rico steffen schmidt & BH. Evolutionary relationship of two ancient protein superfolds. Nat. Chem. Biol. 2014;10:710–715. https://doi.org/10.1038/nchembio.1579.
55. Kolodny R, Nepomnyachiy S, Tawfik DS, Ben-Tal N. Bridging themes: Short protein segments found in different architectures. Mol Biol Evol. 2021;38(6):2191–2208. https://doi.org/10.1093/molbev/msab017.
56. Shortle D. One sequence plus one mutation equals two folds. Proc Natl Acad Sci U S A. 2009;106(50):21011–21012. https://doi.org/10.1073/pnas.0912370107.
57. Meier C, Aricescu AR, Assenberg R, et al. The crystal structure of ORF-9b, a lipid binding protein from the SARS coronavirus. Structure. 2006;14(7):1157–1165. https://doi.org/10.1016/j.str.2006.05.012.
58. Smock RG, Yadid I, Dym O, Clarke J, Tawfik DS. De novo evolutionary emergence of a symmetrical protein is shaped by folding constraints. Cell. 2016;164(3):476–486. https://doi.org/10.1016/j.cell.2015.12.024.
59. Vila JA. Metamorphic proteins in light of Anfinsen's dogma. J Phys Chem. 2020;11(13):4998–4999. https://doi.org/10.1021/acs.jpclett.0c01414.
60. Das M, Chen N, LiWang A, Wang LP. Identification and characterization of metamorphic proteins: Current and future perspectives. Biopolymers. 2021;112(10):e23473. https://doi.org/10.1002/bip.23473.
61. Madhurima K, Nandi B, Sekhar A. Metamorphic proteins: The Janus proteins of structural biology. Open Biol. 2021;11(4):210012. https://doi.org/10.1098/rsob.210012.
62. Dishman AF, Tyler RC, Fox JC, et al. Evolution of fold switching in a metamorphic protein. Science. 2021;371(6524):86–90. https://doi.org/10.1126/science.abd8700.
63. Lella M, Mahalakshmi R. Metamorphic proteins: Emergence of dual protein folds from one primary sequence. Biochemistry. 2017;56(24):2971–2984. https://doi.org/10.1021/acs.biochem.7b00375.
64. Carter CW. Simultaneous codon usage, the origin of the proteome, and the emergence of de-novo proteins. Curr Opin Struct Biol. 2021;68:142–148. https://doi.org/10.1016/j.sbi.2021.01.004.
65. Sabath N, Wagner A, Karlin D. Evolution of viral proteins originated de novo by overprinting. Mol Biol Evol. 2012;29(12):3767–3780. https://doi.org/10.1093/molbev/mss179.
66. Pavesi A, Magiorkinis G, Karlin DG. Viral proteins originated De novo by overprinting can be identified by codon usage: Application to the "gene nursery" of Deltaretroviruses. PLoS Comput Biol. 2013;9(8):e1003162. https://doi.org/10.1371/journal.pcbi.1003162.
67. Bornberg-Bauer E, Schmitz JF. Fact or fiction: Updates on how protein-coding genes might emerge de novo from previously non-coding DNA. F1000Research 2017;6(0). https://doi.org/10.12688/f1000research.10079.1
68. Baalsrud HT, Tørresen OK, Solbakken MH, et al. De novo gene evolution of antifreeze glycoproteins in codfishes revealed by whole genome sequence data. Mol Biol Evol. 2018;35(3):593–606. https://doi.org/10.1093/molbev/msx311.
69. Cai J, Zhao R, Jiang H, Wang W. De novo origination of a new protein-coding gene in Saccharomyces cerevisiae. Genetics. 2008;179(1):487–496. https://doi.org/10.1534/genetics.107.084491.
70. Lange A, Patel PH, Heames B, et al. Structural and functional characterization of a putative de novo gene in drosophila. Nat Commun. 2021;12(1):1–13. https://doi.org/10.1038/s41467-021-21667-6.
71. Szilágyi A, Zhang Y, Závodszky P. Intra-chain 3D segment swapping spawns the evolution of new multidomain protein architectures. J Mol Biol. 2012;415(1):221–235. https://doi.org/10.1016/j.jmb.2011.10.045.
72. Kobe B, Kajava AV. When protein folding is simplified by protein coiling solenoid structures. Trend Biochem Sci. 2000;25(10):509–515.
73. Andrade MA, Perez-Iratxeta C, Ponting CP. Protein repeats: Structures, functions, and evolution. J Struct Biol. 2001;134(2–3):117–131. https://doi.org/10.1006/jsbi.2001.4392.
74. Romero-Romero S, Kordes S, Michel F, Höcker B. Evolution, folding, and design of TIM barrels and related proteins. Curr Opin Struct Biol. 2021;68:94–104. https://doi.org/10.1016/j.sbi.2020.12.007.

75. Söding J, Remmert M, Biegert A. HHrep: De novo protein repeat detection and the origin of TIM barrels. Nucleic Acids Res. 2006;34(suppl_2):137–142. https://doi.org/10.1093/nar/gkl130.
76. Kopec KO, Lupas AN. β-Propeller blades as ancestral peptides in protein evolution. PLoS One. 2013;8(10):e77074. https://doi.org/10.1371/journal.pone.0077074.
77. Lee J, Blaber M. Experimental support for the evolution of symmetric protein architecture from a simple peptide motif. Proc Natl Acad Sci U S A. 2011;108(1):126–130. https://doi.org/10.1073/pnas.1015032108.
78. Fowler DM, Stephany JJ, Fields S. Measuring the activity of protein variants on a large scale using deep mutational scanning. Nat Protoc. 2014;9(9):2267–2284. https://doi.org/10.1038/nprot.2014.153.
79. Heyne M, Shirian J, Cohen I, et al. Climbing up and down binding landscapes through deep mutational scanning of three homologous protein-protein complexes. J Am Chem Soc. 2021;143(41):17261–17275. https://doi.org/10.1021/jacs.1c08707.
80. Newberry RW, Leong JT, Chow ED, Kampmann M, DeGrado WF. Deep mutational scanning reveals the structural basis for α-synuclein activity. Nat Chem Biol. 2020;16(6):653–659. https://doi.org/10.1038/s41589-020-0480-6.
81. Chen JZ, Fowler DM, Tokuriki N. Comprehensive exploration of the translocation, stability and substrate recognition requirements in vim-2 lactamase. Elife. 2020;9:1–31. https://doi.org/10.7554/eLife.56707.
82. Firnberg E, Labonte JW, Gray JJ, Ostermeier M. A comprehensive, high-resolution map of a Gene's fitness landscape. Mol Biol Evol. 2014;31(6):1581–1592. https://doi.org/10.1093/molbev/msu081.
83. Rockah-Shmuel L, Tóth-Petróczy Á, Tawfik DS. Systematic mapping of protein mutational space by prolonged drift reveals the deleterious effects of seemingly neutral mutations. PLoS Comput Biol. 2015;11(8):1–28. https://doi.org/10.1371/journal.pcbi.1004421.
84. Boucher JI, Bolon DNA, Tawfik DS. Quantifying and understanding the fitness effects of protein mutations: Laboratory versus nature. Protein Sci. 2016;25:1219–1226. https://doi.org/10.1002/pro.2928.
85. Wyganowski KT, Kaltenbach M, Tokuriki N. GroEL/ES buffering and compensatory mutations promote protein evolution by stabilizing folding intermediates. J Mol Biol. 2013;425(18):3403–3414. https://doi.org/10.1016/j.jmb.2013.06.028.
86. Bershtein S, Mu W, Serohijos AWR, Zhou J, Shakhnovich EI. Protein quality control acts on folding intermediates to shape the effects of mutations on organismal fitness. Mol Cell. 2013;49(1):133–144. https://doi.org/10.1016/j.molcel.2012.11.004.
87. Kadibalban AS, Bogumil D, Landan G, Dagan T. DnaK-dependent accelerated evolutionary rate in prokaryotes. Genome Biol Evol. 2016;8(5):1590–1599. https://doi.org/10.1093/gbe/evw102.
88. Young DL, Fields S. The role of functional data in interpreting the effects of genetic variation. Mol Biol Cell. 2015;26(22):3904–3908. https://doi.org/10.1091/mbc.E15-03-0153.
89. De Vos MGJ, Dawid A, Sunderlikova V, Tans SJ. Breaking evolutionary constraint with a tradeoff ratchet. Proc Natl Acad Sci U S A. 2015;112(48):14906–14911. https://doi.org/10.1073/pnas.1510282112.
90. Noda-García L, Davidi D, Korenblum E, et al. Chance and pleiotropy dominate genetic diversity in complex bacterial environments. Nat Microbiol. 2019;4(7):1221–1230. https://doi.org/10.1038/s41564-019-0412-y.
91. Echave J, Spielman SJ, Wilke CO. Causes of evolutionary rate variation among protein sites. Nat Rev Genet. 2016;17(2):109–121. https://doi.org/10.1038/nrg.2015.18.
92. Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. Nature. 2006;444(7121):929–932. https://doi.org/10.1038/nature05385.
93. Yang JR, Liao BY, Zhuang SM, Zhang J. Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. Proc Natl Acad Sci U S A. 2012;109(14):5158–5159. https://doi.org/10.1073/pnas.1117408109.
94. Tóth-Petróczy Á, Tawfik DS. Slow protein evolutionary rates are dictated by surface - core association. Proc Natl Acad Sci U S A. 2011;108(27):11151–11156. https://doi.org/10.1073/pnas.1015994108.
95. Tokuriki N, Jackson CJ, Afriat-Jurnou L, Wyganowski KT, Tang R, Tawfik DS. Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. Nat Commun. 2012;3:1257. https://doi.org/10.1038/ncomms2246.
96. Gade M, Tan LL, Damry AM, et al. Substrate dynamics contribute to enzymatic specificity in human and bacterial methionine Adenosyltransferases. JACS. 2021;1:2349–2360. https://doi.org/10.1021/jacsau.1c00464.
97. Ben-David M, Soskine M, Dubovetskyi A, et al. Enzyme evolution: An epistatic ratchet versus a smooth reversible transition. Mol Biol Evol. 2020;37(4):1133–1147. https://doi.org/10.1093/molbev/msz298.
98. Clarkson CS, Temple HJ, Miles A. The genomics of insecticide resistance: Insights from recent studies in African malaria vectors. Curr Opin Insect Sci. 2018;27:111–115. https://doi.org/10.1016/j.cois.2018.05.017.
99. Menozzi P, Shi MA, Lougarre A, Tang ZH, Fournier D. Mutations of acetylcholinesterase which confer insecticide resistance in Drosophila melanogaster populations. BMC Evol Biol. 2004;4:1–7. https://doi.org/10.1186/1471-2148-4-4.
100. Hartley CJ, Newcomb RD, Russell RJ, et al. Amplification of DNA from preserved specimens shows blowflies were preadapted for the rapid evolution of insecticide resistance. Proc Natl Acad Sci U S A. 2006;103(23):8757–8762. https://doi.org/10.1073/pnas.0509590103.
101. MacLean RC, Torres-Barceló C, Moxon R. Evaluating evolutionary models of stress-induced mutagenesis in bacteria. Nat Rev Genet. 2013;14(3):221–227. https://doi.org/10.1038/nrg3415.
102. Zhou K, Aertsen A, Michiels CW. The role of variable DNA tandem repeats in bacterial adaptation. FEMS Microbiol Rev. 2014;38(1):119–141. https://doi.org/10.1111/1574-6976.12036.
103. Jee J, Rasouly A, Shamovsky I, et al. Rates and mechanisms of bacterial mutagenesis from maximum-depth sequencing. Nature. 2016;534(7609):693–696. https://doi.org/10.1038/nature18313.
104. Morreall J, Kim A, Liu Y, Degtyareva N, Weiss B, Doetsch PW. Evidence for Retromutagenesis as a mechanism

for adaptive mutation in Escherichia coli. PLoS Genet. 2015; 11(8):1–12. https://doi.org/10.1371/journal.pgen.1005477.
105. Sankar TS, Wastuwidyaningtyas BD, Dong Y, Lewis SA, Wang JD. The nature of mutations induced by replication-transcription collisions. Nature. 2016;535(7610):178–181. https://doi.org/10.1038/nature18316.
106. Carelli FN, Hayakawa T, Go Y, Imai H, Warnefors M, Kaessmann H. The life history of retrocopies illuminates the evolution of new mammalian genes. Genome Res. 2016;26(3): 301–314. https://doi.org/10.1101/gr.198473.115.
107. Slack A, Thornton PC, Magner DB, Rosenberg SM, Hastings PJ. On the mechanism of gene amplification induced under stress in Escherichia coli. PLoS Genet. 2006;2(4):385–398. https://doi.org/10.1371/journal.pgen.0020048.
108. Wannarat W, Wannarat W, Motoyama S, Masuda K, Kawamura F, Inaoka T. Tetracycline tolerance mediated by gene amplification in Bacillus subtilis. Microbiol. 2014;160: 2474–2480. https://doi.org/10.1099/mic.0.081505-0.
109. Adler M, Anjum M, Berg OG, Andersson DI, Sandegren L. High fitness costs and instability of gene duplications reduce rates of evolution of new genes by duplication-divergence mechanisms. Mol Biol Evol. 2014;31(6):1526–1535. https://doi.org/10.1093/molbev/msu111.
110. Näsvall J, Sun L, Roth JR, Andersson DI. Real-time evolution of new genes by innovation, amplification, and divergence. Science. 2012;338(6105):384–387. https://doi.org/10.1126/science.1226521.
111. Graves CJ, Ros VID, Stevenson B, Sniegowski PD, Brisson D. Natural selection promotes antigenic Evolvability. PLoS Pathog. 2013;9(11):e1003766. https://doi.org/10.1371/journal.ppat.1003766.
112. Rotem E, Loinger A, Ronin I, et al. Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence. Proc Natl Acad Sci U S A. 2010;107(28): 12541–12546. https://doi.org/10.1073/pnas.1004333107.
113. Garcia-Bernardo J, Dunlop MJ. Phenotypic diversity using bimodal and unimodal expression of stress response proteins. Biophys J. 2016;110(10):2278–2287. https://doi.org/10.1016/j.bpj.2016.04.012.
114. Ribas de Pouplana L, Santos MAS, Zhu JH, Farabaugh PJ, Javid B. Protein mistranslation: Friend or foe? Trends Biochem Sci. 2014;39(8):355–362. https://doi.org/10.1016/j.tibs.2014.06.002.
115. Gordon AJE, Satory D, Halliday JA, Herman C. Lost in transcription: Transient errors in information transfer. Curr Opin Microbiol. 2015;24:80–87. https://doi.org/10.1016/j.mib.2015.01.010.
116. Verheijen BM, van Leeuwen FW. Commentary: The landscape of transcription errors in eukaryotic cells. Front Genet. 2017;8. https://www.science.org/doi/10.1126/sciadv.1701484116.
117. Ling J, O'Donoghue P, Söll D. Genetic code flexibility in microorganisms: Novel mechanisms and impact on physiology. Nat Rev Microbiol. 2015;13(11):707–721. https://doi.org/10.1038/nrmicro3568.
118. Mordret E, Dahan O, Asraf O, et al. Systematic detection of amino acid substitutions in proteomes reveals mechanistic basis of ribosome errors and selection for translation Fidelity. Mol Cell. 2019;75(3):427–441.e5. https://doi.org/10.1016/j.molcel.2019.06.041.
119. Bratulic S, Gerber F, Wagner A. Mistranslation drives the evolution of robustness in TEM-1 β-lactamase. Proc Natl Acad Sci U S A. 2015;112(41):12758–12763. https://doi.org/10.1073/pnas.1510071112.
120. Yanagida H, Gispan A, Kadouri N, et al. The evolutionary potential of phenotypic mutations. PLoS Genet. 2015;11(8):1–20. https://doi.org/10.1371/journal.pgen.1005445.
121. Ruiz-Orera J, Messeguer X, Subirana JA, Alba MM. Long non-coding RNAs as a source of new peptides. Elife. 2014;3:1–24. https://doi.org/10.7554/eLife.03523.
122. Pauli A, Valen E, Schier AF. Identifying (non-)coding RNAs and small peptides: Challenges and opportunities. Bioessays. 2015;37(1):103–112. https://doi.org/10.1002/bies.201400103.
123. Saghatelian A, Couso JP. Discovery and characterization of smORF-encoded bioactive polypeptides. Nat Chem Biol. 2015; 11(12):909–916. https://doi.org/10.1038/nchembio.1964.
124. Chen S, Krinsky BH, Long M. New genes as drivers of phenotypic evolution. Nat Rev Genet. 2013;14(9):645–660. https://doi.org/10.1038/nrg3521.
125. McLysaght A, Guerzoni D. New genes from non-coding sequence: The role of de novo protein-coding genes in eukaryotic evolutionary innovation. Philos Trans R Soc B Biol Sci. 2015;370(1678):20140332. https://doi.org/10.1098/rstb.2014.0332.
126. Halfmann R, Jarosz DF, Jones SK, Chang A, Lancaster AK, Lindquist S. Prions are a common mechanism for phenotypic inheritance in wild yeasts. Nature. 2012;482(7385):363–368. https://doi.org/10.1038/nature10875.
127. Hornung G, Bar-Ziv R, Rosin D, Tokuriki N, Tawfik DS, Oren M, Barkai N. Noise-mean relationship in mutated promoters. Genome Res 2012;22(12):2409–2417. https://doi.org/10.1101/gr.139378.112
128. Raser JM, O'Shea EK. Control of stochasticity in eukaryotic gene expression. Science. 2004;304(5678):1811–1814. https://doi.org/10.1126/science.1098641.
129. Levin BR, Rozen DE. Opinion - Non-inherited antibiotic resistance. Nature Reviews Microbiology. 2006;4:556–562.
130. Allan Drummond D, Wilke CO. The evolutionary consequences of erroneous protein synthesis. Nat Rev Genet. 2009; 10(10):715–724. https://doi.org/10.1038/nrg2662.
131. Rockah-Shmuel L, Tóth-Petróczy Á, Sela A, Wurtzel O, Sorek R, Tawfik DS. Correlated occurrence and bypass of frame-shifting insertion-deletions (InDels) to give functional proteins. PLoS Genet. 2013;9(10):e1003882. https://doi.org/10.1371/journal.pgen.1003882.
132. van der Woude MW. Phase variation: How to create and coordinate population diversity. Curr Opin Microbiol. 2011;14(2): 205–211. https://doi.org/10.1016/j.mib.2011.01.002.
133. Atkins JF, Loughran G, Bhatt PR, Firth AE, Baranov PV. Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. Nucleic Acids Res. 2016;44(15):7007–7078. https://doi.org/10.1093/nar/gkw530.
134. Yi X, Dean AM. Phenotypic plasticity as an adaptation to a functional trade-off. Elife. 2016;5:1–12. https://doi.org/10.7554/elife.19307.

135. Miller SR, Longley R, Hutchins PR, Bauersachs T. Cellular innovation of the cyanobacterial heterocyst by the adaptive loss of plasticity. Curr Biol. 2020;30(2):344–350.e4. https://doi.org/10.1016/j.cub.2019.11.056.
136. Corl A, Bi K, Luke C, et al. The genetic basis of adaptation following plastic changes in coloration in a novel environment. Curr Biol. 2018;28(18):2970, e7–2977. https://doi.org/10.1016/j.cub.2018.06.075.
137. Li A, Li L, Zhang Z, Li S, Wang W, Guo X. Noncoding variation and transcriptional plasticity promote thermal adaptation in oysters by altering energy metabolism. Mol Biol Evol. 2021;38(11):5144–5155. https://doi.org/10.1093/molbev/msab241.
138. Samhita L, Raval PK, Agashe D. Global mistranslation increases cell survival under stress in Escherichia coli. PLoS Genet. 2020;16(3):1–21. https://doi.org/10.1371/journal.pgen.1008654.
139. Zheng J, Guo N, Wagner A. Mistranslation reduces mutation load in evolving proteins through negative epistasis with DNA mutations. Mol Biol Evol. 2021;38(11):4792–4804. https://doi.org/10.1093/molbev/msab206.
140. Copley SD. Moonlighting is mainstream: Paradigm adjustment required. Bioessays. 2012;34(7):578–588. https://doi.org/10.1002/bies.201100191.
141. Rauwerdink A, Lunzer M, Devamani T, et al. Evolution of a catalytic mechanism. Mol Biol Evol. 2016;33(4):971–979. https://doi.org/10.1093/molbev/msv338.
142. Amitai G, Gupta RD, Tawfik DS. Latent evolutionary potentials under the neutral mutational drift of an enzyme. HFSP J. 2007;1(1):67–78. https://doi.org/10.2976/1.2739115.
143. Bershtein S, Tawfik DS. Ohno's model revisited: Measuring the frequency of potentially adaptive mutations under various mutational drifts. Mol Biol Evol. 2008;25(11):2311–2318. https://doi.org/10.1093/molbev/msn174.
144. Sabater-Muñoz B, Prats-Escriche M, Montagud-Martínez R, et al. Fitness trade-offs determine the role of the molecular chaperonin GroEL in buffering mutations. Mol Biol Evol. 2015;32(10):2681–2693. https://doi.org/10.1093/molbev/msv144.
145. Liang H, Lin YS, Li WH. Fast evolution of core promoters in primate genomes. Mol Biol Evol. 2008;25(6):1239–1244. https://doi.org/10.1093/molbev/msn072.
146. Hittinger CT, Carroll SB. Gene duplication and the adaptive evolution of a classic genetic switch. Nature. 2007;449(7163):677–681. https://doi.org/10.1038/nature06151.
147. Gu Z, Nicolae D, Lu HHS, Li WH. Rapid divergence in expression between duplicate genes inferred from microarray data. Trends Genet. 2002;18(12):609–613. https://doi.org/10.1016/S0168-9525(02)02837-8.
148. Makova KD, Li WH. Divergence in the spatial pattern of gene expression between human duplicate genes. Genome Res. 2003;13(7):1638–1645. https://doi.org/10.1101/gr.1133803.
149. Lavy T, Yanagida H, Tawfik DS. Gal3 binds Gal80 tighter than Gal1 indicating adaptive protein changes following duplication. Mol Biol Evol. 2016;33(2):472–477. https://doi.org/10.1093/molbev/msv240.
150. Noda-Garcia L, Romero Romero ML, Longo LM, Kolodkin-Gal I, Tawfik DS. Bacilli glutamate dehydrogenases diverged via coevolution of transcription and enzyme regulation. EMBO Rep. 2017;18(7):1139–1149. https://doi.org/10.15252/embr.201743990.
151. Lynch M. The evolution of multimeric protein assemblages. Mol Biol Evol. 2012;29(5):1353–1366. https://doi.org/10.1093/molbev/msr300.
152. Pillai AS, Chandler SA, Liu Y, et al. Origin of complexity in haemoglobin evolution. Nature. 2020;581(7809):480–485. https://doi.org/10.1038/s41586-020-2292-y.
153. Garcia-Seisdedos H, Empereur-Mot C, Elad N, Levy ED. Proteins evolve on the edge of supramolecular self-assembly. Nature. 2017;548(7666):244–247. https://doi.org/10.1038/nature23320.
154. Marques AC, Vinckenbosch N, Brawand D, Kaessmann H. Functional diversification of duplicate genes through subcellular adaptation of encoded proteins. Genome Biol. 2008;9(3):1–12. https://doi.org/10.1186/gb-2008-9-3-r54.
155. Liu SL, Pan AQ, Adams KL. Protein subcellular relocalization of duplicated genes in Arabidopsis. Genome Biol Evol. 2014;6(9):2501–2515. https://doi.org/10.1093/gbe/evu191.
156. Nilsson AI, Koskiniemi S, Eriksson S, Kugelberg E, Hinton JCD, Andersson DI. Bacterial genome size reduction by experimental evolution. Proc Natl Acad Sci U S A. 2005;102(34):12112–12116. https://doi.org/10.1073/pnas.0503654102.
157. Jensen RA. Enzyme recruitment in evolution of new function. Annu Rev Microbiol. 1976;30:409–425. https://doi.org/10.1146/annurev.mi.30.100176.002205.
158. Weng JK. The evolutionary paths towards complexity: A metabolic perspective. New Phytol. 2014;201(4):1141–1149. https://doi.org/10.1111/nph.12416.
159. Notebaart RA, Szappanos B, Kintses B, et al. Network-level architecture and the evolutionary potential of underground metabolism. Proc Natl Acad Sci USA. 2014;111(32):11762–11767. https://doi.org/10.1073/pnas.1406102111.
160. Caetano-Anollés K, Caetano-Anollés G. Structural Phylogenomics reveals gradual evolutionary replacement of abiotic chemistries by protein enzymes in purine metabolism. PLoS One. 2013;8(3):e59300. https://doi.org/10.1371/journal.pone.0059300.
161. Keller MA, Turchyn AV, Ralser M. Non-enzymatic glycolysis and pentose phosphate pathway-like reactions in a plausible Archean Ocean. Mol Syst Biol. 2014;10(4):1–12. https://doi.org/10.1002/msb.20145228.
162. Laurino P, Tawfik DS. Spontaneous emergence of S-Adenosylmethionine and the evolution of methylation. Angew Chemie. 2017;129(1):349–351. https://doi.org/10.1002/ange.201609615.
163. Lazcano A, Miller SL. On the origin of metabolic pathways. J Mol Evol. 1999;49(4):424–431. https://doi.org/10.1007/PL00006565.
164. Cornish-Bowden A, Cárdenas ML. Life before LUCA. J Theor Biol. 2017;434:68–74. https://doi.org/10.1016/j.jtbi.2017.05.023.
165. Piette BMAG, Heddle JG. A Peptide–Nucleic Acid Replicator Origin for Life. Trends Ecol Evol. 2020;35(5):397–406. https://doi.org/10.1016/j.tree.2020.01.001.
166. Chatterjee S, Yadav S. The origin of prebiotic information system in the peptide/RNA world: A simulation model of the evolution of translation and the genetic code. Life. 2019;9(1). https://doi.org/10.3390/life9010025.
167. Rode BM. Peptides and the origin of life. Peptides. 1999;20(6):773–786. https://doi.org/10.1016/S0196-9781(99)00062-5.

168. Ikehara K. Possible steps to the emergence of life: The [GADV]-protein world hypothesis. Chem Rec. 2005;5(2):107–118. https://doi.org/10.1002/tcr.20037.
169. Fried SD, Fujishima K, Makarov M, Cherepashuk I, Hlouchova K. Peptides before and during the nucleotide world: An origins story emphasizing cooperation between proteins and nucleic acids. J R Soc Interface. 2022;19(187):20210641. https://doi.org/10.1098/rsif.2021.0641.
170. Maury CPJ. Amyloid and the origin of life: Self-replicating catalytic amyloids as prebiotic informational and protometabolic entities. Cell Mol Life Sci. 2018;75(9):1499–1507. https://doi.org/10.1007/s00018-018-2797-9.
171. Friedmann MP, Torbeev V, Zelenay V, Sobol A, Greenwald J, Riek R. Towards prebiotic catalytic amyloids using high throughput screening. PLoS One. 2015;10(12):1–16. https://doi.org/10.1371/journal.pone.0143948.
172. Romero Romero ML, Yang F, Lin YR, et al. Simple yet functional phosphate-loop proteins. Proc Natl Acad Sci U S A. 2018;115(51):E11943–E11950. https://doi.org/10.1073/pnas.1812400115.
173. Longo LM, Petrovic D, Kamerlin SCL, Tawfik DS. Short and simple sequences favored the emergence of N-helix phospholigand binding sites in the first enzymes. Proc Natl Acad Sci U S A. 2020;117(10):5310–5318. https://doi.org/10.1073/pnas.1911742117.
174. Medvedev KE, Kinch LN, Schaeffer RD, Grishin NV. Functional Analysis of Rossmann-like Domains Reveals Convergent Evolution of Topology and Reaction Pathways. Plos Comput Biol. 2019;15. https://doi.org/10.1371/journal.pcbi.1007569.
175. Söding J, Lupas AN. More than the sum of their parts: On the evolution of proteins from peptides. Bioessays. 2003;25(9):837–846. https://doi.org/10.1002/bies.10321.
176. Ferruz N, Lobos F, Lemm D, Identification and analysis of natural building blocks for evolution-guided fragment-based protein design. S Toledo-Patino, J A Farías-Rico, S Schmidt, B Höcker, J Mol Biol 2020;432(13):3898–3914. https://doi.org/10.1016/j.jmb.2020.04.013
177. Alva V, Söding J, Lupas AN. A vocabulary of ancient peptides at the origin of folded proteins. Elife. 2015;4:1–19. https://doi.org/10.7554/elife.09410.
178. van Oss SB, Carvunis AR. De novo gene birth. PLoS Genet. 2019;15(5):1–23. https://doi.org/10.1371/journal.pgen.1008160.
179. Wilson BA, Foy SG, Neme R, Masel J. Young genes are highly disordered as predicted by the preadaptation hypothesis of de novo gene birth. Nat Ecol Evol. 2017;1(6):1–19. https://doi.org/10.1038/s41559-017-0146.
180. Bungard D, Copple JS, Yan J, et al. Foldability of a natural De novo evolved protein. Structure. 2017;25(11):1687, e4–1696. https://doi.org/10.1016/j.str.2017.09.006.
181. Tawfik DS. Accuracy-rate tradeoffs: How do enzymes meet demands of selectivity and catalytic efficiency? Curr Opin Chem Biol. 2014;21:73–80. https://doi.org/10.1016/j.cbpa.2014.05.008.
182. Innan H, Kondrashov F. The evolution of gene duplications: Classifying and distinguishing between models. Nat Rev Genet. 2010;11(2):97–108. https://doi.org/10.1038/nrg2689.
183. Noda-García L, Camacho-Zarco AR, Medina-Ruíz S, et al. Evolution of substrate specificity in a recipient's enzyme following horizontal gene transfer. Mol Biol Evol. 2013;30(9):2024–2034. https://doi.org/10.1093/molbev/mst115.
184. Dellus-Gur E, Elias M, Caselli E, et al. Negative epistasis and evolvability in TEM-1 β-lactamase - the thin line between an enzyme's conformational freedom and disorder. J Mol Biol. 2015;427(14):2396–2409. https://doi.org/10.1016/j.jmb.2015.05.011.
185. Starr TN, Thornton JW. Epistasis in protein evolution. Protein Sci. 2016;25:1204–1218. https://doi.org/10.1002/pro.2897.
186. Miton CM, Tokuriki N. How mutational epistasis impairs predictability in protein evolution and design. Protein Sci. 2016;25:1260–1272. https://doi.org/10.1002/pro.2876.
187. Harms MJ, Thornton JW. Historical contingency and its biophysical basis in glucocorticoid receptor evolution. Nature. 2014;512(7513):203–207. https://doi.org/10.1038/nature13410.
188. 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;4:1–26. https://doi.org/10.7554/eLife.07864.
189. Dellus-Gur E, Toth-Petroczy A, Elias M, Tawfik DS. What makes a protein fold amenable to functional innovation? Fold polarity and stability trade-offs. J Mol Biol. 2013;425(14):2609–2621. https://doi.org/10.1016/j.jmb.2013.03.033.
190. Payne JL, Wagner A. The robustness and evolvability of transcription factor binding sites. Science. 2014;343(6173):875–877. https://doi.org/10.1126/science.1249046.
191. Sikosek T, Chan HS. Biophysics of protein evolution and evolutionary protein biophysics. J R Soc Interface. 2014;11(100):20140419. https://doi.org/10.1098/rsif.2014.0419.
192. Liskova V, Bednar D, Prudnikova T, et al. Balancing the stability-activity trade-off by fine-tuning dehalogenase access tunnels. ChemCatChem. 2015;7(4):648–659. https://doi.org/10.1002/cctc.201402792.
193. Newton MS, Arcus VL, Patrick WM. Rapid bursts and slow declines: On the possible evolutionary trajectories of enzymes. J R Soc Interface. 2015;12(107):20150036. https://doi.org/10.1098/rsif.2015.0036.

How to cite this article: Jayaraman V, Toledo-Patiño S, Noda-García L, Laurino P. Mechanisms of protein evolution. Protein Science. 2022;31(7):e4362. https://doi.org/10.1002/pro.4362