Linking molecular evolution to molecular grafting

Conan K. Wang*[ID] and David J. Craik[ID]

From the Institute for Molecular Bioscience and Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science, The University of Queensland, Brisbane, Queensland, Australia

Edited by Ursula Jakob

Molecular grafting is a strategy for the engineering of molecular scaffolds into new functional agents, such as next-generation therapeutics. Despite its wide use, studies so far have focused almost exclusively on demonstrating its utility rather than understanding the factors that lead to either poor or successful grafting outcomes. Here, we examine protein evolution and identify parallels between the natural process of protein functional diversification and the artificial process of molecular grafting. We discuss features of natural proteins that are correlated to innovability—the capacity to acquire new functions—and describe their implications to molecular grafting scaffolds. Disulfide-rich peptides are used as exemplars because they are particularly promising scaffolds onto which new functions can be grafted. This article provides a perspective on why some scaffolds are more suitable for grafting than others, identifying opportunities on how molecular grafting might be improved.

Proteins with their diversified activities constitute the primary functional units of biology, working together in coordinated interaction networks that are essential for life. However, their dysfunction or dysregulation is the cause of many diseases. In pursuit of more efficacious therapeutics, drug discovery has moved from being a serendipitous pursuit to one aimed at precise control of molecular function and the deliberate targeting of specific proteins involved in disease pathogenesis. One approach to the design of such targeted drug leads is molecular grafting—the transplantation of foreign functional amino acids onto a proteinaceous scaffold to create a stable molecule with therapeutic function (1). Examples where molecular grafting has been successfully used to design potent drug leads and protein scaffolds that have shown promise in clinical trials have been reviewed (2–4). The process of molecular grafting shares striking parallels with the natural molecular evolution of proteins, yet the two have not previously been linked.

Here, we explore molecular grafting from a new perspective by linking it to concepts derived from protein evolution. To establish this connection, we begin this article by comparing the two processes in more detail, first from the viewpoint of molecular grafting and then from the standpoint of evolution. From there, we discuss studies on protein evolution, particularly those that explore factors governing functional diversification. The term innovability—the capacity to acquire new

functions—earlier introduced by Tawfik *et al.* is used here to recognize related and seminal studies of protein diversification and enzyme engineering (5). The term is used to additionally bring the focal point onto evolutionary-derived factors that could also be important in molecular grafting. We then discuss these factors with reference to disulfide-rich peptide scaffolds to explore new ways of evaluating why certain scaffolds have been difficult to use in molecular grafting and thus risk becoming extinct, whereas others are inherently destined to thrive. We conclude by speculating on how evolutionary information might be used to change not only how we think about the process of molecular grafting but also how it is carried out in the future.

Comparison of molecular grafting to molecular evolution

Molecular grafting and "evolution" of the scaffold

Molecular grafting has similarities to its namesakes in horticulture or surgery. Although it has broad applications in biotechnology, we focus here on its use in drug discovery. In that application, molecular grafting involves taking a bioactive pharmacophore and grafting it onto a scaffold—the intended outcome being a therapeutic agent bestowed with desired traits from both the "insert" and the scaffold (Fig. 1). The pharmacophore insert could be a peptide chain that inhibits a target protein activity, or a noncontiguous set of hotspot residues rendered from the target protein–protein interaction interface. The scaffolds can be as small as 50-mer peptides or as large and complex as multidomain monoclonal antibodies, although the term molecular grafting is more frequently associated with the former type of scaffold, and particularly with disulfide-rich peptides (1).

A recurring theme in molecular grafting is the selection of scaffolds for their naturally evolved properties, such as disulfide-rich peptides for their stability (1) and drug-like properties (6), single protein domains for their compact modular size (3, 7), or monoclonal antibodies for their role in protein binding (8). The use of scaffolds with known structure and well-characterized function underpins much of the power of molecular grafting in drug design because it reduces the complexity of a problem that is essentially a search through an astronomically large sequence space. Molecular grafting can be viewed as a more tractable approach than *de novo* design, notwithstanding the recent significant progress in the latter (9–11). In fact, through evolution, the generation of new

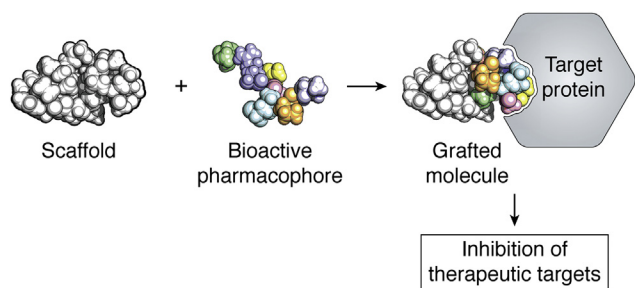* For correspondence: Conan K. Wang, c.wang@imb.uq.edu.au.

Figure 1. Molecular grafting of bioactive residues onto a scaffold. The scaffold, shown in *white*, acts as a molecular canvas, upon which bioactive residues are grafted, resulting in a variant of the scaffold that now has a new function. Molecular grafting has been applied to drug design to develop therapeutic leads that inhibit a target protein.

protein domains that are functional has commonly occurred through the diversification of existing protein domains and rarely through the creation of *de novo* folds ([12–14]). This observation points to the effectiveness of designing new functional molecules by re-purposing existing protein entities.

Both molecular grafting and molecular evolution involve the modification of a progenitor protein to endow it with new function. One difference between the two processes is the definition of fitness, with economic factors, for example, having an influence in drug discovery, but not in the same context as in evolution. Another difference is that molecular grafting is conducted over much shorter time frames and with smaller population sizes. For instance, a recombinant library method (such as phage, bacterial, yeast display) will typically have a library size of $\sim 10^{12}$, whereas a structure-guided rational method will be limited by throughput of $< 10^3$ candidates in a typical academic laboratory. These two methods represent the two extremes on the spectrum of approaches that have been employed for molecular grafting.

Models of evolution and their relation to molecular grafting

Diversification of protein function from an ancestral domain involves the evolutionary processes of duplication, mutation, and selection. Ohno's classical evolution model ([15], [16]) speculates duplication to be a neutral event, with copies drifting under no selection, thereby accumulating mutations that might or might not affect protein function. A protein with new function subsequently emerges from this evolved library of variants through selection. Ohno's model has been questioned for its evolutionary accuracy because duplication has an energetic cost and is therefore unlikely to be exempt from selection ([17]). Furthermore, mutations are, on average, deleterious and hence believed to accumulate under purifying selection to remove those that cause protein misfolding or loss of function instead of being under no selection. Modern evolutionary models account for these additional selective pressures ([17]).

Like molecular evolution, molecular grafting can be viewed as involving duplication, mutation, and selection. Ohno's model can be used to describe molecular grafting by recombinant library methods, in which duplicated copies of a

scaffold (rather than a domain) are randomly mutated in an unbiased manner. Mutants having the desired activity or biopharmaceutical properties are selected for using functional assays. A drawback of Ohno's model in this context (or for adaptation) is that it is a relatively inefficient process. Many misfolded and inactive variants of the scaffold need to be screened (or maintained) before a functional one is found, making creation of a new function a costly exercise. If the library from which function is selected could first be subjected to purifying selection to remove deleterious copies, as accounted for in modern evolutionary models, then the discovery of new function would be a more likely event. Despite this insight from evolution, deliberate removal of misfolded variants from the initial recombinant library remains a nontrivial task. Nevertheless, the important implication from evolutionary models is that the challenge of using recombinant display is not only the size of the library but also the way in which it is designed and subsequently enriched or optimized. Approaches that combine recombinant libraries to increase throughput with intelligent structure-guided design of libraries, preferentially comprising folded members to increase discovery efficiency, should therefore be of substantial value.

Factors that drive evolution and the diversification of function

Differential rates of protein evolution

The frequency of diversification has increased through evolutionary time, as supported by the correlation between growth of protein domain families and organism complexity ([18], [19]). The proteomes of eukaryotes comprise heavily duplicated proteins, more so than in prokaryotes. Expansion of protein domain families in metazoa and vertebrates is associated with functions required for multicellularity, such as cellular signaling and regulation, and elaborate extracellular sensing, such as immunity. These functions are mediated through interaction networks, implying proteins enriched through duplication in multicellular organisms have a fitness advantage for protein binding. Additionally, these proteins have been allowed to diversify to evolve new functions, suggesting they are also advantageous for acquiring new activities ([20]). These inferences support the proposition that frequently duplicated proteins are collectively an excellent source of scaffolds for molecular grafting because the aim is to design molecules that bind protein targets and redress dysregulated networks.

The apparently biased duplication of certain protein domains could have other explanations, including different domains being subject to different evolutionary times and/or the effect of chance ([20–22]). Domains that were initially favored by chance would have represented a larger proportion of available proteins from which subsequent proteins evolve, establishing an avalanche-like effect. Different proteins are also under different selective pressures, which affect their rate of evolution. For example, proteins that are abundant or perform multiple important functions are under higher selective pressure because mutations would have a more pronounced effect

compared with those that are scarce or relatively nonessential [17]. In drug design, scaffolds are mutated and selected for function, but other biopharmaceutical properties, not typically a priority for nature, such as metabolic stability and cost of manufacture, should also be considered.

Defining innovability and its determinants

Notwithstanding the complexity underlying differential rates of evolution, if the frequency of duplication of a protein domain is simply a manifestation of functional versatility, then an important question that has been raised is as follows: what features of those domains are conducive to that phenomenon [23]? In molecular grafting, these features are relevant because they could help guide the selection of scaffolds to increase the success rate of obtaining the desired bioproduct. In attempts to formalize these features, several terms have been proposed to classify protein domains and folds, including "innovability" defined above [20]. Innovability is thought to be a prerequisite for evolvability, the capacity of a protein domain to change along evolutionary time [20]. So too is robustness, the ability to tolerate mutations while maintaining function and structure [20]. Another term, designability, has previously been proposed to describe the number of sequences that adopt the desired fold as the ground state and govern how readily a

protein fold is selected for in nature [24]. In molecular grafting, we are primarily concerned with the design of *new* function while maintaining structure of the scaffold to be achieved over a much shorter timescale than that associated with evolution. Therefore, we choose to borrow the concept of innovability henceforth when discussing molecular grafting.

Several seemingly contrasting biophysical features have been proposed to affect innovability during evolution. On the one hand, configurational stability is thought to be a key attribute [5]. Ordered structured proteins, such as Rossman folds and TIM barrels, are buffered with many intramolecular contacts available to compensate against the potentially destabilizing effects of new mutations, granting those proteins high innovability [5, 25, 26], as illustrated in Figure 2A. In an apparently conflicting observation, highly disordered proteins can also exhibit high evolutionary rates [27], potentially because mutations have little influence on their ability to function and therefore have high innovability, as depicted in Figure 2B. Furthermore, conformational plasticity and dynamics, rather than rigidity, is correlated with functional promiscuity, another feature thought to enable innovability [28–31]. These disparate concepts have raised the question of whether there exists a structural order–disorder paradox [20].
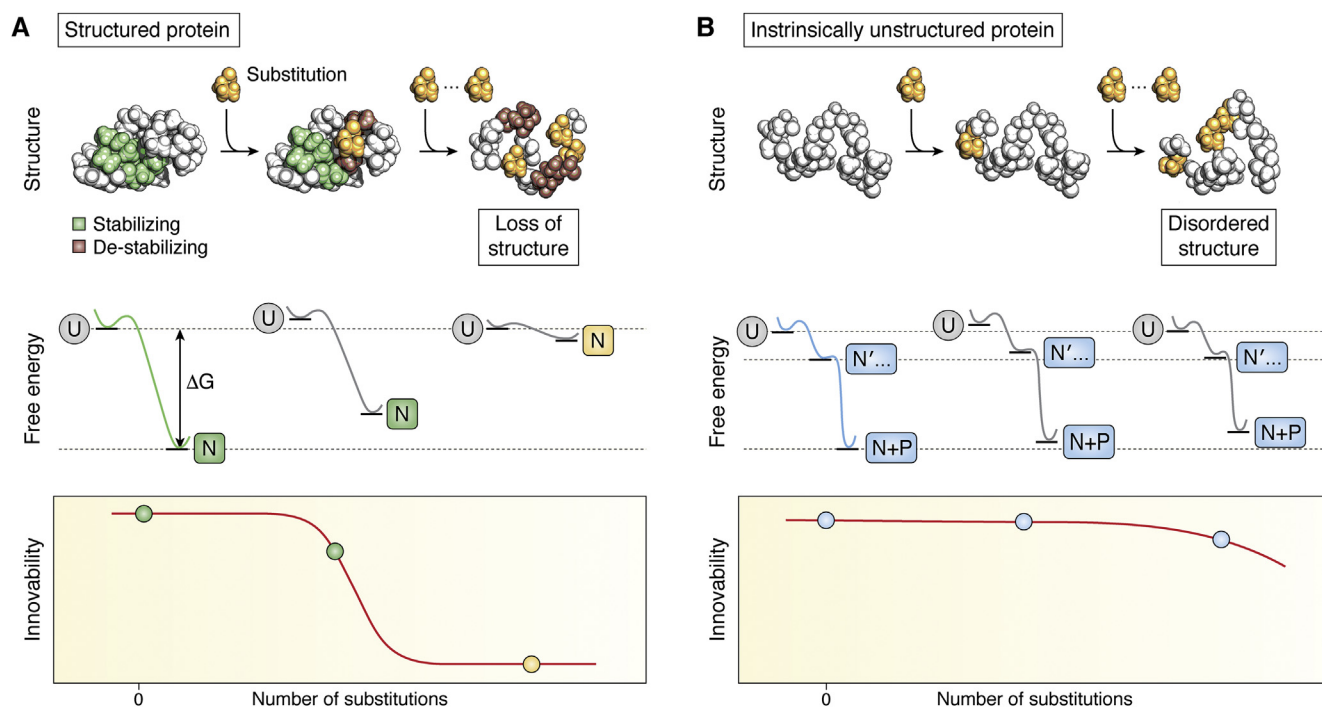

Figure 2. Structure, disorder, and innovability. Both highly structured and disordered proteins have exhibited high innovability or the capacity to acquire new functions, leading to the paradoxical conclusion that both structure and disorder are possible prerequisites for acquiring new functions. *A*, a general example for a structured protein. A substitution (in *orange*) to a highly structured protein that causes an unfavorable energetic change to the native state would introduce de-stabilizing contacts with preexisting residues (*maroon*) but would be tolerated by the structure because of the network of stabilizing residue contacts (*green*), as shown in the *upper panel*. In other words, the energetic change introduced by the substitution would be small relative to the free energy difference (ΔG) between the unfolded (U) and native states (N), as shown in the *middle panel*. These proteins have many stabilizing interactions that withstand the destabilizing effects of many substitutions before loss of structural stability and, by inference, innovability, as shown in the lower panel. Consequently, configurational stability is thought to be important for innovability. *B*, a general example for intrinsically unstructured proteins. These proteins have no definite folded state for unstructured proteins, existing as a mixture of conformations (N' …) that becomes more defined upon binding to a protein target (N + P). Mutations are likely to be tolerated in unstructured proteins, making them highly innovable.

One way to resolve the paradox is to consider that nature utilizes both structured and disordered proteins that can evolve new functions. Disordered proteins tend to be associated with transient low-affinity interactions, which are beneficial in signaling networks to allow a rapid response to changing and unpredictable stimuli (32). By comparison, structured proteins are associated with long-lived high-affinity interactions that would be beneficial for precise and committed interactions. For these interactions that are often essential and require specificity, the restrictions that structure places on innovability would be an advantage because it helps resist evolutionary change that would, on average, have a negative effect on fitness.

The compensation between order and disorder and innovability could well be context dependent. For instance, if folding into a well-formed structure is a prerequisite for function, then determinants of configurational stability would naturally be associated with innovability. In the opposing context, where structure is not vital, then disordered proteins would have higher innovability than structured proteins because they contain a larger proportion of sites that can be mutated without structural effect. In drug design, rigidity is often a constraint on the scaffold from the outset because it is associated with many desired pharmaceutical properties that increase efficacy and reduce off-target effects, such as enzymatic stability and high binding affinity and specificity.

Innovability at the residue level

Rates of evolution can vary for individual sites within a protein (33, 34). Sites that evolve slowly are generally under selective pressure to maintain structure and function (33). Biophysical factors, such as solvent accessibility, contact density, and flexibility have been correlated to evolutionary rates, although they have shown limited predictive accuracy (34). An important insight has come from mutagenesis and evolutionary analysis of enzymes, which examined the role of residues in stabilizing the scaffold structure and on activity (35). The results have led to the concept of active site–scaffold polarity, which defines the degree of separation of residues mediating function (*e.g.*, active-site residues) from those that maintain fold (*e.g.*, core residues) (20, 23, 36). Enzyme folds that have high polarity are proposed to have high innovability. For example, the TIM barrel exemplifies a highly innovable fold, whereas DHRF exemplifies a noninnovable fold based on the polarity concept (see ref 35 for an illustration of active-site polarity). Identification of residues that mediate function can sometimes be obscure as, although they are typically built around the active site, they can also occur at distant surfaces that are connected to the protein core through physically contiguous contact networks, referred to as sectors (37). These internal contact networks appear to be ubiquitous in proteins and are related to conserved functional activities. Adapting the idea of polarity to molecular grafting, sites that should be chosen for mutation should be those that are decoupled to the structure of the scaffold. Within the set of decoupled residues, disordered regions would be more innovable but might lead to lower affinity interactions, whereas rigid regions would be less innovable but might result in more specific binding to the target.

Re-thinking disulfide-rich scaffolds in molecular grafting

Disulfide-rich scaffolds and their evolutionary diversity

Peptides comprising disulfide bonds are highly abundant and diverse in nature. Interestingly, a very limited correlation has been observed between the sequence and structures of these types of peptides (38), supporting their potential for high innovability. We focus on peptides having fewer than 50 amino acids because they have beneficial biopharmaceutical properties based on their small size, such as increased tissue penetration and low potential for immunogenicity compared with protein scaffolds. A recent survey of the Protein Data Bank structural database identified a set of disulfide-rich peptides of varying cysteine content and connectivity (39). Peptides with three disulfide bonds forming a I-IV, II-V, III-VI connectivity are the most common. A large proportion of these form the inhibitor cystine knot (ICK) topology, in which two disulfides (I-IV and II-V) form a ring through which the third disulfide (III-VI) threads, as shown in Figure 3A. Peptides that comprise the ICK structural motif are also known as knottins (40) and are highly dispersed through evolution, with examples found in fungi (41), plants (42), and animals (43, 44), as illustrated in Figure 3A. For example, the fungal tomato pathogen *Cladosporium fulvum* produces AVR9, a 28-amino acid peptide (45); the Australian funnel-web spider *Hadronyche infensa* produces Hi1a, a 75-residue peptide (46); and the potato produces a 39-amino acid carboxypeptidase inhibitor (47). The structural and functional diversity of the ICK motif show that it has acquired different mutations and activity over evolutionary time, suggesting that peptides with this motif are highly innovable. Indeed, they have attracted wide interest as therapeutic modalities (48–52). For example, analogs of the scorpion peptide chlorotoxin (53) and the engineered R01-MG peptide (54) have undergone clinical trials as imaging tools to guide surgical oncology, and an analog of the cyclotide kalata B1 is entering trials as an immunomodulatory agent (55).

The cyclotides are a family of ICK-containing peptides that have additionally evolved a macrocyclic backbone (56). The sequences that project from the ICK core are referred to as loops, with each loop connecting two successive Cys residues: the first loop (loop 1) joins I to II and the last completing the cycle between VI and I. Cyclotides have so far been found in numerous plant species, with a single host plant capable of producing hundreds of different cyclotides (Fig. 3A) (57). They function in plant defense, with some acting by targeting gut membranes of insect predators and others by inhibiting digestive enzymes (58). Their protective functions provide an evolutionary explanation for their distribution and abundance. Their gene structures, which can contain multiple copies of the same or different cyclotide precursors (Fig. 3B), is evidence of duplication and drift through evolution (59–61), mirroring the process of protein diversification described in evolutionary models.
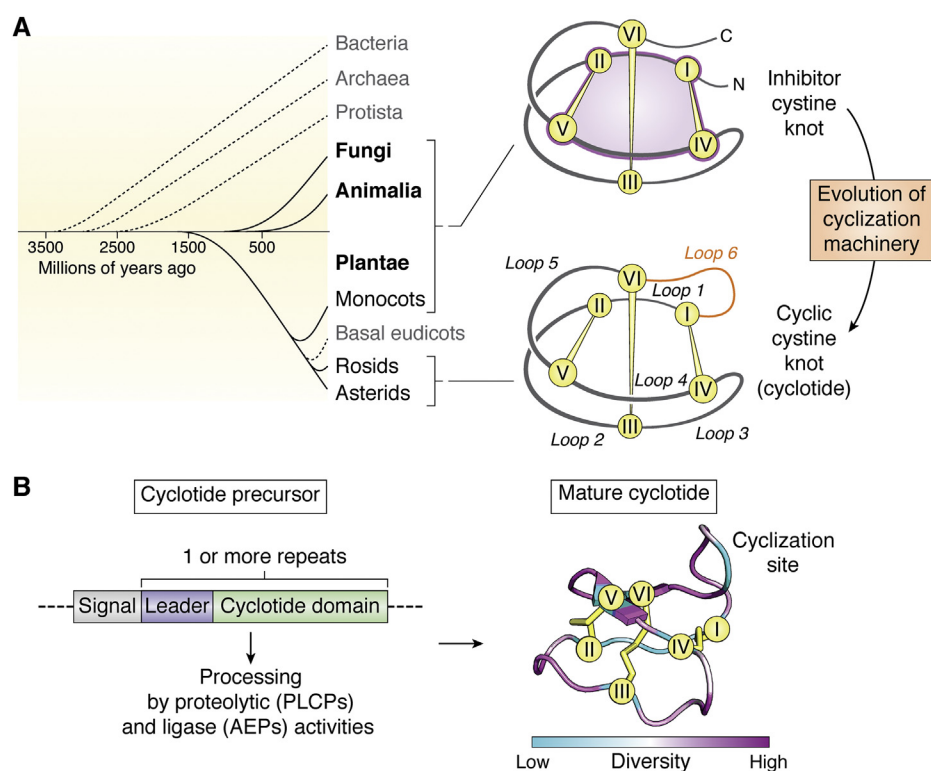
Figure 3. Evolution of the inhibitor cystine knot motif as a natural scaffold. *A,* the inhibitor cystine knot (ICK) is a structural motif that is widely distributed in nature and found in many different organisms of varying evolutionary age. Kingdoms and clades in which ICK-containing proteins or peptides have been found so far are highlighted in *black font*, whereas kingdoms and clades that do not have ICK-proteins and peptides are in *gray font*. Kingdom names are *bolded* to distinguish them from clades. Evidence for the existence of the ICK in different organisms is based on entries in the KNOTTIN database (40) and independent experimental validation. The ICK motif comprises three disulfide bonds linked in a I-IV, II-V, and III-VI connectivity as shown to the top right. The first two disulfide bonds bridge the interleaving sequences between the cysteine residues I and II and IV and V, forming a ring (in *purple*) that encircles the third disulfide. One family of ICK-containing peptides have a cyclic cystine knot motif and are called cyclotides. They are distinguished by also having a backbone sequence linking the sixth cysteine residue back to the first (in *orange*). The sequences between successive Cys residues are referred to as loops, which are labeled 1–6. Cyclotides have been found in the Rosids and Asterids clades. *B,* cyclotides are ribosomally produced from gene precursors that encode one or more cyclotides, which is evidence of evolution involving gene or domain duplication. Processing of the gene precursors into mature cyclotides involves proteolytic and ligase activities of papain-like cysteine proteases and asparaginyl endopeptidases. Based on the reported cyclotides sequences, cyclotides exhibit large sequence diversity in their loop sequences. Conservation of residues in loop 6 is associated with the biosynthetic pathway.

ICK-containing peptides have been widely used as molecular grafting scaffolds, with cyclotides represented in over two dozen studies reported so far (1, 62). The most frequently used cyclotide scaffolds are kalata B1, discovered from *Oldenlandia affinis*, from the coffee plant family, and MCoTI-II from *Momordica cochinchinensis* from the squash plant family. CyBase, a database of cyclotide sequence and structures (63), shows cyclotides have high sequence diversity which, when mapped onto the prototypical structure of a cyclotide, provides evidence for their potential as a molecular grafting scaffold for tolerating new mutations, as illustrated in Figure 3B. Conserved residues in the cyclotide scaffold that are related to structure (*i.e.,* Cys residues) would have low innovability. The conserved Asp/Asn residue in loop 6, however, is a result of an enzyme-mediated biosynthetic mechanism (64–66) that can be easily bypassed using chemical ligation strategies during artificial synthesis (58). It is important to note that the evolutionary reason for the conservation of specific residues in loop six are different to those that govern innovability.

Configurational stability of the scaffold structure

High melting temperatures are diagnostic of stable scaffolds and are exhibited in many reported molecular grafting scaffolds, including those demonstrating high clinical potential so far, such as the monobody/adnectin scaffold (2). Of the scaffolds used in molecular grafting studies, disulfide-rich peptides arguably represent optimal solutions. The cyclotide scaffold, for example, contains three interlocked disulfides that covalently reinforce the core against thermal or chemical denaturation. Its rigid core has been shown to compensate against Ala substitutions to all loop residues (*i.e.,* non-Cys residues), with the mutations having no apparent effect on overall structure of the prototypical cyclotide, kalata B1 (67). Applying concepts of evolution, the capacity of the loops to tolerate mutations indicates they are decoupled from the structural residues. Loop 6 appears to be the most decoupled of all cyclotide loops because it is not present in many other ICK-containing peptides (40) and therefore not essential for structure. Indeed, an acyclic mutant of kalata B1, in which loop

6 is disconnected, retains the cyclotide fold (68, 69). Interestingly, the cyclotide fold is also tolerant to acyclic permutation (68, 69), which is also a feature of thermostable proteins with a robust and densely packed core (70).

Not all peptides belonging to disulfide-rich scaffolds characterized by high sequence diversity have rigid folds and are highly innovable; *e.g.*, the peptide based on the EGF(A) domain from the LDL receptor, which is involved in cholesterol metabolism by binding to the serum protein PCSK9. The EGF(A) domain belongs to the EGF-like family, which encompasses many disulfide-rich domains found in humans (71). Unlike the cyclotide kalata B1, it has poor tolerance to Ala substitutions to non-Cys residues, with the mutations destabilizing the overall fold and favoring highly disordered states despite retaining the disulfide bonds (72). The mutational analysis suggests almost all residues are coupled to structure, implying the peptide has low innovability as a scaffold. Why are so many residues coupled to the structure in this peptide but not in others? The answer might lie in its specialized function as a calcium sensor, which requires it to rapidly exchange between conformations in response to fluctuating calcium levels (72). It appears evolution has fine-tuned the peptide and all its constituent residues to collectively perform a specific purpose, potentially pushing the peptide into a "dead-end" in the innovability landscape. This example shows that it is important to delve deeper into the evolutionary history of a peptide to understand its potential innovability, and the current day snapshot of the diversity of the family to which it belongs can be misleading.

Folding into the scaffold structure

The discussion above focuses on the stability of the structure once formed, but of course that formation process cannot be taken for granted. Indeed, the capacity of a scaffold to fold into its native structure and deliver its payload is an important determinant of its innovability. Folding is generally considered to be dependent on sequence only, in accordance with Alfinsen's dogma (73), and described by funneled energy landscapes directed toward the lowest energy state, *e.g.*, the native fold for natural proteins. Mutations that reduce yield of the native structure reconfigure the folding landscape, creating obstructive energetic barriers that disrupt intramolecular contacts required for productive folding. Therefore, in molecular grafting, it is preferred to have scaffolds with robust folding landscapes stabilized by productive contacts during folding that resist being derailed by mutations. In choosing regions of a scaffold to modify, one should consider regions decoupled to productive folding pathways, as regions decoupled to structure are likely to have high innovability. However, interresidue contacts involved in folding might be cryptic because they might exist only transiently during folding and disappear once the final structure has formed. For the cyclotide scaffold, Ala substitutions have vastly different effects on folding yield, despite having no significant effect on the integrity of the final structure (67), as shown in Figure 4. Mutagenesis studies are

therefore useful for identifying regions of a scaffold coupled to folding. Experimental and theoretical approaches to probing protein folding has been reviewed previously (74).

According to the principle of minimal frustration (75), evolution optimizes the sequences of natural proteins to fold efficiently. This principle is supported by the observation that homologous proteins have comparable folding rates despite having different melting temperatures because of different unfolding rates (76). It could infer that scaffolds associated with high diversity have robust folding pathways. However, the level to which folding has been optimized can be dependent on evolutionary time. The family of α-conotoxin peptides found in the venom of cone snails are two disulfide-containing peptides characterized by highly diverse sequences, but native peptides are well known to easily adopt different disulfide bond isomers. It is possible that conotoxins have not been exposed to sufficient evolutionary time to be optimized to fold into a single active conformation, or cone snails have not yet evolved elaborate assistive folding mechanisms compared with more complex organisms. Interestingly, despite the structure-function dogma, different disulfide isomers of conotoxins can still be functionally active (77–79), which might suggest their folding plasticity has been selected to offer an adaptive advantage. Nevertheless, as folding is important to the innovability of a peptide scaffold but not straightforward to understand based on an analysis of sequence diversity, directed experiments are useful to help define the level of innovability of a chosen scaffold.

Future directions

The aim of this article is to establish a connection between molecular evolution and molecular grafting, from which we hope to stimulate wider investigations into the use of evolutionary information to drive the design of next-generation therapeutics. How can we learn from evolution to select better scaffolds for molecular grafting? Can evolutionary processes be used to improve ways in which we modify scaffolds? Are studies of evolution also useful beyond molecular design of function? Can they be applied to help translate a grafted scaffold into an approved therapeutic?

Insightful uses of evolutionary information could lead to scaffolds that are more suitable for molecular grafting (Fig. 5). The naïve approach is to select peptides or proteins that are associated with a sequence-diverse fold as scaffolds. However, this type of evolutionary analysis does not account for evolutionary processes or phylogenetic relationships. It provides only a starting pool of potential scaffolds, rather than a conclusive assessment on the innovability of a selected member. Methods to identify/design those with high thermal stability might be more fruitful in identifying a single starting scaffold (5), and we propose they should have wider applications in molecular grafting. For example, mutating positions to the consensus amino acid could be used to design starting scaffolds. The method has been proposed as a means of increasing thermostability based on the hypothesis that the consensus amino acid contributes more to stability than
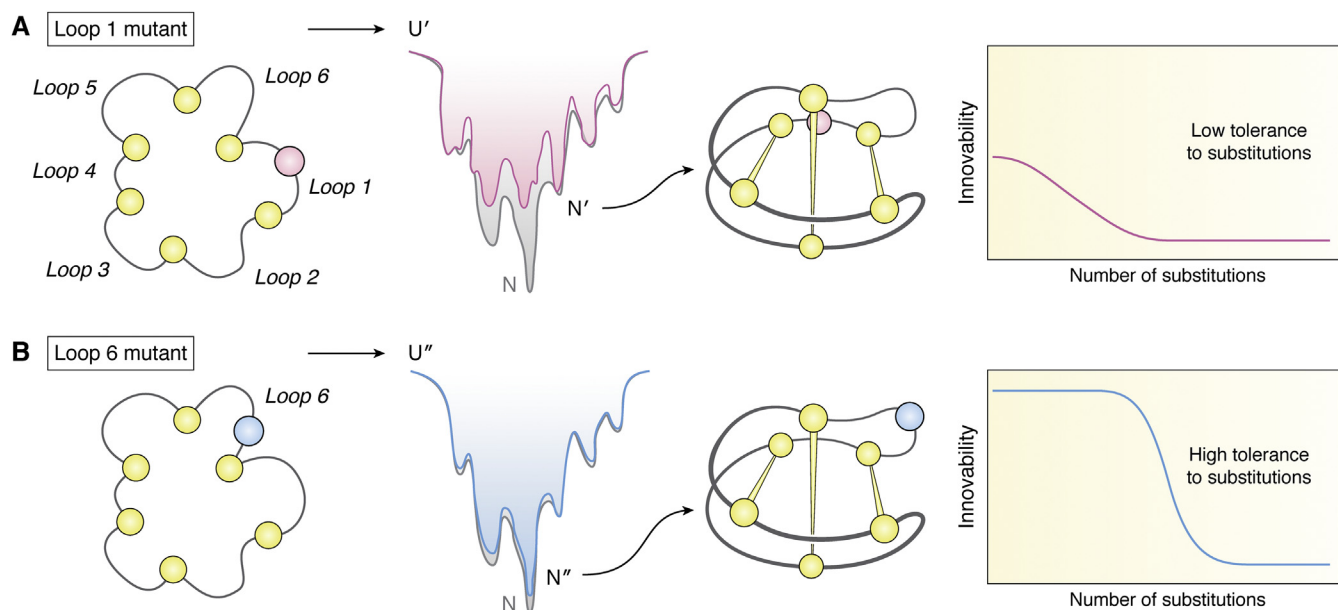
Figure 4. Innovable regions within a scaffold. A scaffold will have regions within it that have higher innovability than others depending on its level of decoupling from stability of the scaffold. Stability is affected by folding and unfolding rates. In this figure, we demonstrate the variation of innovability among sites within a scaffold associated with folding by using the cyclotide scaffold as the example. The cyclotide scaffold is depicted as a string with the characteristic six Cys residues as *yellow circles*. Overall, a native cyclotide scaffold might have high innovability and be able to accommodate many substitutions. *A*, however, some regions are coupled to folding. For example, when a substitution is introduced into loop 1 (*red circle*), the folding landscape is distorted (*pink*), making folding into the native structure (N′) less favorable than that achievable by the parent peptide (N, *gray* folding landscape). Therefore, loop 1 has low starting innovability that declines rapidly with each additional substitution. *B*, in comparison, other regions (*e.g.*, loop 6) are decoupled from folding, and so substitutions in these regions (*blue circle*) have no impact on the folding landscape. The native structure of scaffold analogs (N″) are similarly favored as the native state of the parent scaffold (N). Therefore, these regions have high innovability and high tolerance to substitutions.

nonconsensus residues (80, 81). Ancestral sequence reconstruction is another method for predicting a highly stable scaffold (82, 83). This alternative method is based on the hypothesis that ancestral proteins have higher thermal stability than those observed today, with the notion that environmental temperatures were higher in the Precambrian era as one possible explanation, although the origins of the stability and trends across evolutionary time are still under debate (84). Not directly related to stability is the method of identifying hub sequences for use as starting scaffolds (85). These sequences are highly connected to related sequences through mutational space. That is, compared with poorly connected sequences, hub sequences require fewer mutational changes to describe more of the observed sequences. These techniques have proven successful in protein design and enzyme engineering efforts but have yet to be fully explored from a molecular grafting perspective. Looking to the future, improvements to the accuracy of these methods in predicting optimal scaffolds and a thorough understanding of their utility in molecular grafting would be valuable.

The search for innovable scaffolds should be followed by an understanding of how a selected scaffold can be best modified. The importance of having regions decoupled to structure and function for innovability argue for the need for detailed understanding of the biophysical role of individual residues and therefore the importance of advancing experimental techniques that enable those investigations to be conducted rapidly. It also points to the value in examining links between

evolution and protein physics (86), *e.g.*, between statistical physics and population genetics, to build mechanistic explanatory models that could be simulated to understand protein mutational drivers and predict yet unknown determinants of innovability. There are opportunities for applying co-evolutionary analyses, which have demonstrated potential for inferring function (87), predicting functional dynamics (88) and tracing protein specificity and affinity (89), to reimagine processes used in molecular grafting to design molecules that target a specific protein.

We recognize there are other considerations beyond scaffold selection in drug development (Fig. 5). These considerations could be related to pharmacodynamic properties (*e.g.*, target selectivity and specificity), pharmacokinetic properties (*e.g.*, route of administration, cytoplasmic delivery efficiency for cytosolic targets), or developability (*e.g.*, cost of production). A therapeutic biomolecule often needs to meet many criteria to have a broad impact on healthcare. Some scaffolds might inherently be more suitable regardless of their evolutionary history. For example, smaller scaffolds, such as peptides, have higher tumor tissue penetration and will be preferred for applications where that property is a priority. Although peptide and protein-based therapeutics can be viewed as safe because they are composed of amino acids and therefore breakdown into products already found in the human body, their potential immunogenicity is a constant concern. It is therefore interesting that sequences designed by consensus engineering and ancestral reconstruction have
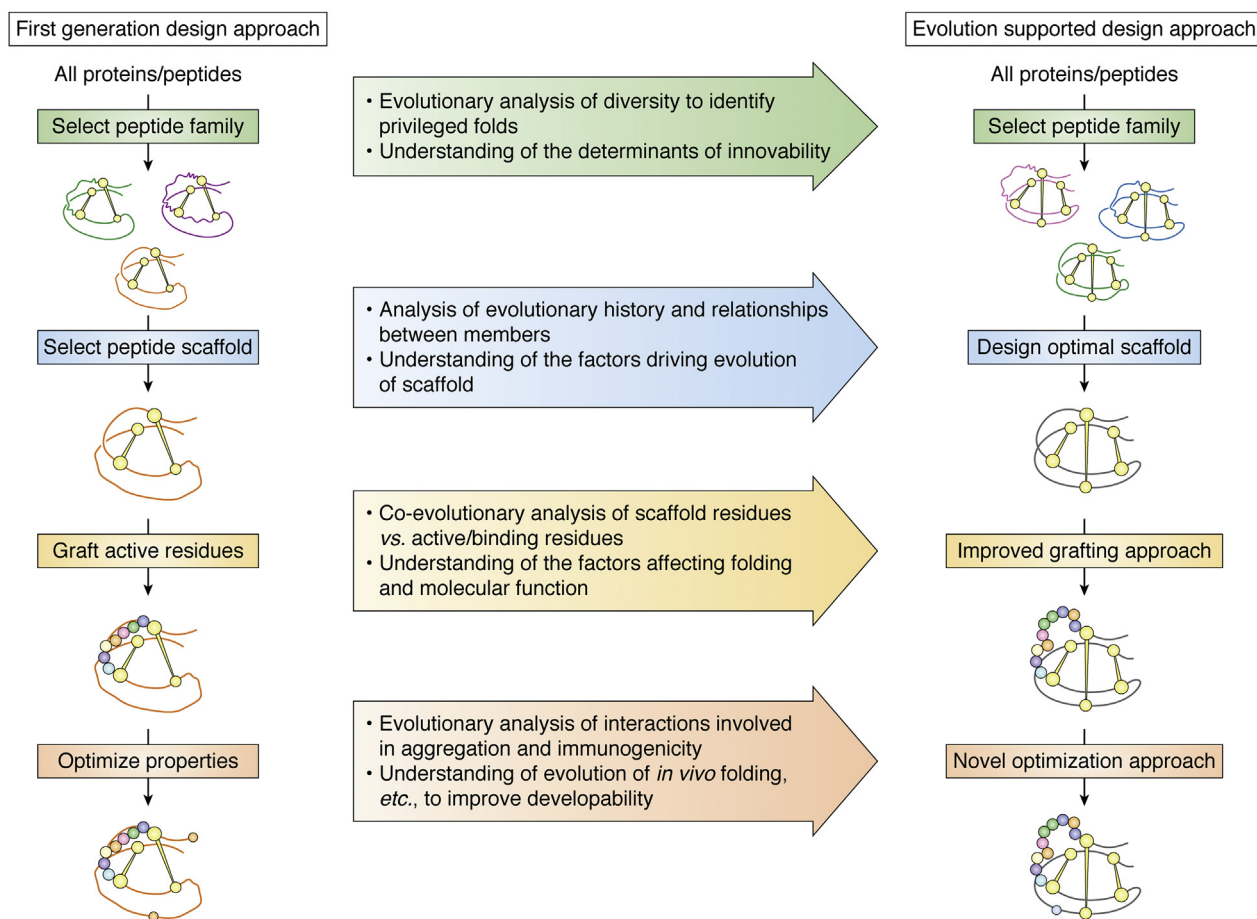
Figure 5. Use of evolutionary information to improve the process of molecular grafting. The early conceptualization of molecular grafting, referred to as the "first-generation approach", is represented in the vertical flow diagram on the *left*. It involves the initial identification of a naturally occurring family of peptides chosen for their potential as scaffolds, from which an individual member is then selected based on a desired property, such as natively high expression or stability. A bioactive peptide sequence (shown as a string of colored circles representing amino acids) is then grafted onto this lead scaffold, resulting in a grafted peptide that typically requires optimization before it can become a therapeutic agent (schematically illustrated by the two-disulfide exemplar grafted and optimized peptide at the *bottom left*). Disulfide bonds are shown in *yellow*, and *orange-colored circles* are used to illustrate residue mutations during optimization. Although some aspects of evolution are intrinsically considered during this first-generation approach, we propose there to be many opportunities to incorporate evolutionary information to improve the process, leading to what we here call an "evolution supported approach". Such an approach, shown in the flow diagram on the right, might lead to the more successful design of drug-like molecules. Several types of evolutionary analyses and the resultant understanding of specific aspects of protein diversity, co-evolution, and impacts of aggregation and immunogenicity are noted in the text *boxes* to indicated how these factors can be built into the evolution supported approach.

potential for lower immunogenicity (90), although the reasons for that observation in a few cases so far are still unclear. A biological therapeutic also needs to resist aggregation to maintain desired efficacy. An understanding of how biological systems have evolved to avoid unwanted aggregation might be helpful here. For example, a study has shown that low sequence identity between consecutive domains in a tandem array safeguards against misfolding and aggregation (91), which is useful knowledge for design of multivalent designer proteins. We mention these studies that link evolution to desired biopharmaceutical traits because they raise the intriguing thought that evolutionary information will be able to contribute more to the design of therapeutics beyond the identification of scaffolds.

In conclusion, we have provided a perspective on molecular grafting by comparing it to evolution. We highlighted the concept of innovability and suggest an understanding of its

determinants is valuable for the selection of scaffolds that have the best chance of leading to a successful grafting outcome. Beyond these investigations, evolutionary processes might have broader value in inspiring new approaches to design and develop therapeutics.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviation—The abbreviations used are: ICK, inhibitor cystine knot.

References

1. Wang, C. K., and Craik, D. J. (2018) Designing macrocyclic disulfide-rich peptides for biotechnological applications. *Nat. Chem. Biol.* **14**, 417–427
2. Vazquez-Lombardi, R., Phan, T. G., Zimmermann, C., Lowe, D., Jermutus, L., and Christ, D. (2015) Challenges and opportunities for non-antibody scaffold drugs. *Drug Discov. Today* **20**, 1271–1283
3. Gebauer, M., and Skerra, A. (2020) Engineered protein scaffolds as next-generation therapeutics. *Annu. Rev. Pharmacol. Toxicol.* **60**, 391–415
4. Gilbreth, R. N., and Koide, S. (2012) Structural insights for engineering binding proteins based on non-antibody scaffolds. *Curr. Opin. Struct. Biol.* **22**, 413–420
5. Trudeau, D. L., and Tawfik, D. S. (2019) Protein engineers turned evolutionists-the quest for the optimal starting point. *Curr. Opin. Biotechnol.* **60**, 46–52
6. Crook, Z. R., Sevilla, G. P., Friend, D., Brusniak, M. Y., Bandaranayake, A. D., Clarke, M., Gewe, M., Mhyre, A. J., Baker, D., Strong, R. K., Bradley, P., and Olson, J. M. (2017) Mammalian display screening of diverse cystine-dense peptides for difficult to drug targets. *Nat. Commun.* **8**, 2244
7. Simeon, R., and Chen, Z. (2018) In vitro-engineered non-antibody protein therapeutics. *Protein Cell* **9**, 3–14
8. Chiu, M. L., Goulet, D. R., Teplyakov, A., and Gilliland, G. L. (2019) Antibody structure and function: The Basis for engineering therapeutics. *Antibodies (Basel)* **8**, 55
9. Korendovych, I. V., and DeGrado, W. F. (2020) De novo protein design, a retrospective. *Q. Rev. Biophys.* **53**, e3
10. Bhardwaj, G., Mulligan, V. K., Bahl, C. D., Gilmore, J. M., Harvey, P. J., Cheneval, O., Buchko, G. W., Pulavarti, S. V., Kaas, Q., Eletsky, A., Huang, P. S., Johnsen, W. A., Greisen, P. J., Rocklin, G. J., Song, Y., *et al.* (2016) Accurate de novo design of hyperstable constrained peptides. *Nature* **538**, 329–335
11. Dawson, W. M., Rhys, G. G., and Woolfson, D. N. (2019) Towards functional de novo designed proteins. *Curr. Opin. Chem. Biol.* **52**, 102–111
12. Kolodny, R., Pereyaslavets, L., Samson, A. O., and Levitt, M. (2013) On the universe of protein folds. *Annu. Rev. Biophys.* **42**, 559–582
13. Tautz, D., and Domazet-Loso, T. (2011) The evolutionary origin of orphan genes. *Nat. Rev. Genet.* **12**, 692–702
14. Bornberg-Bauer, E., Huylmans, A. K., and Sikosek, T. (2010) How do new proteins arise? *Curr. Opin. Struct. Biol.* **20**, 390–396
15. Ohno, S. (1970) *Evolution by Gene Duplication*, Springer, NY
16. Kimura, M., and Ohta, T. (1974) On some principles governing molecular evolution. *Proc. Natl. Acad. Sci. U. S. A.* **71**, 2848–2852
17. Soskine, M., and Tawfik, D. S. (2010) Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.* **11**, 572–582
18. Bagowski, C. P., Bruins, W., and Te Velthuis, A. J. (2010) The nature of protein domain evolution: Shaping the interaction network. *Curr. Genomics* **11**, 368–376
19. Forslund, S. K., Kaduk, M., and Sonnhammer, E. L. L. (2019) Evolution of protein domain Architectures. *Methods Mol. Biol.* **1910**, 469–504
20. Toth-Petroczy, A., and Tawfik, D. S. (2014) The robustness and innovability of protein folds. *Curr. Opin. Struct. Biol.* **26**, 131–138
21. Choi, S. S., and Hannenhalli, S. (2013) Three independent determinants of protein evolutionary rate. *J. Mol. Evol.* **76**, 98–111
22. Toll-Riera, M., Bostick, D., Alba, M. M., and Plotkin, J. B. (2012) Structure and age jointly influence rates of protein evolution. *Plos Comput. Biol.* **8**, e1002542
23. Toth-Petroczy, A., and Tawfik, D. S. (2011) Slow protein evolutionary rates are dictated by surface-core association. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11151–11156
24. Li, H., Helling, R., Tang, C., and Wingreen, N. (1996) Emergence of preferred structures in a simple model of protein folding. *Science* **273**, 666–669
25. Tokuriki, N., and Tawfik, D. S. (2009) Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **19**, 596–604
26. England, J. L., and Shakhnovich, E. I. (2003) Structural determinant of protein designability. *Phys. Rev. Lett.* **90**, 218101
27. Brown, C. J., Johnson, A. K., Dunker, A. K., and Daughdrill, G. W. (2011) Evolution and disorder. *Curr. Opin. Struct. Biol.* **21**, 441–446
28. Campitelli, P., Modi, T., Kumar, S., and Ozkan, S. B. (2020) The role of conformational dynamics and Allostery in Modulating protein evolution. *Annu. Rev. Biophys.* **49**, 267–288
29. Crean, R. M., Gardner, J. M., and Kamerlin, S. C. L. (2020) Harnessing conformational plasticity to generate designer enzymes. *J. Am. Chem. Soc.* **142**, 11324–11342
30. Pabis, A., Risso, V. A., Sanchez-Ruiz, J. M., and Kamerlin, S. C. (2018) Cooperativity and flexibility in enzyme evolution. *Curr. Opin. Struct. Biol.* **48**, 83–92
31. Campbell, E., Kaltenbach, M., Correy, G. J., Carr, P. D., Porebski, B. T., Livingstone, E. K., Afriat-Jurnou, L., Buckle, A. M., Weik, M., Hollfelder, F., Tokuriki, N., and Jackson, C. J. (2016) The role of protein dynamics in the evolution of new enzyme function. *Nat. Chem. Biol.* **12**, 944–950
32. Perkins, J. R., Diboun, I., Dessailly, B. H., Lees, J. G., and Orengo, C. (2010) Transient protein-protein interactions: Structural, functional, and network properties. *Structure* **18**, 1233–1243
33. Echave, J., and Wilke, C. O. (2017) Biophysical models of protein evolution: Understanding the Patterns of evolutionary sequence Divergence. *Annu. Rev. Biophys.* **46**, 85–103
34. Echave, J., Spielman, S. J., and Wilke, C. O. (2016) Causes of evolutionary rate variation among protein sites. *Nat. Rev. Genet.* **17**, 109–121
35. Elias, M., and Tawfik, D. S. (2013) What makes a protein fold amenable to functional innovation? fold polarity and stability trade-offs. *J. Mol. Biol.* **425**, 2609–2621
36. Dellus-Gur, E., Toth-Petroczy, A., Elias, M., and Tawfik, D. S. (2013) What makes a protein fold amenable to functional innovation? fold polarity and stability trade-offs. *J. Mol. Biol.* **425**, 2609–2621
37. McLaughlin, R. N., Jr., Poelwijk, F. J., Raman, A., Gosal, W. S., and Ranganathan, R. (2012) The spatial architecture of protein function and adaptation. *Nature* **491**, 138–142
38. Correnti, C. E., Gewe, M. M., Mehlin, C., Bandaranayake, A. D., Johnsen, W. A., Rupert, P. B., Brusniak, M. Y., Clarke, M., Burke, S. E., De Van Der Schueren, W., Pilat, K., Turnbaugh, S. M., May, D., Watson, A., Chan, M. K., *et al.* (2018) Screening, large-scale production and structure-based classification of cystine-dense peptides. *Nat. Struct. Mol. Biol.* **25**, 270–278
39. Yin, H., Craik, D. J., and Wang, C. K. (2019) Anchor residues guide form and function in grafted peptides. *Angew. Chem. Int. Ed. Engl.* **58**, 7652–7656
40. Postic, G., Gracy, J., Perin, C., Chiche, L., and Gelly, J. C. (2018) KNOTTIN: The database of inhibitor cystine knot scaffold after 10 years, toward a systematic structure modeling. *Nucleic Acids Res.* **46**, D454–D458
41. Zhao, J., Yuan, S., Gao, B., and Zhu, S. (2018) Molecular diversity of fungal inhibitor cystine knot peptides evolved by domain repeat and fusion. *FEMS Microbiol. Lett.* **365**, 1–8
42. Molesini, B., Treggiari, D., Dalbeni, A., Minuz, P., and Pandolfini, T. (2017) Plant cystine-knot peptides: Pharmacological perspectives. *Br. J. Clin. Pharmacol.* **83**, 63–70
43. Pineda, S. S., Chin, Y. K., Undheim, E. A. B., Senff, S., Mobli, M., Dauly, C., Escoubas, P., Nicholson, G. M., Kaas, Q., Guo, S., Herzig, V., Mattick, J. S., and King, G. F. (2020) Structural venomics reveals evolution of a complex venom by duplication and diversification of an ancient peptide-encoding gene. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 11399–11408
44. Jin, A. H., Muttenthaler, M., Dutertre, S., Himaya, S. W. A., Kaas, Q., Craik, D. J., Lewis, R. J., and Alewood, P. F. (2019) Conotoxins: Chemistry and biology. *Chem. Rev.* **119**, 11510–11549
45. Vervoort, J., van den Hooven, H. W., Berg, A., Vossen, P., Vogelsang, R., Joosten, M. H., and de Wit, P. J. (1997) The race-specific elicitor AVR9 of the tomato pathogen Cladosporium fulvum: A cystine knot protein. Sequence-specific 1H NMR assignments, secondary structure and global fold of the protein. *FEBS Lett.* **404**, 153–158
46. Chassagnon, I. R., McCarthy, C. A., Chin, Y. K., Pineda, S. S., Keramidas, A., Mobli, M., Pham, V., De Silva, T. M., Lynch, J. W., Widdop, R. E.,

Rash, L. D., and King, G. F. (2017) Potent neuroprotection after stroke afforded by a double-knot spider-venom peptide that inhibits acid-sensing ion channel 1a. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 3750–3755

47. Rees, D. C., and Lipscomb, W. N. (1980) Structure of potato inhibitor complex of carboxypeptidase A at 5.5-A resolution. *Proc. Natl. Acad. Sci. U. S. A.* **77**, 277–280

48. Kintzing, J. R., and Cochran, J. R. (2016) Engineered knottin peptides as diagnostics, therapeutics, and drug delivery vehicles. *Curr. Opin. Chem. Biol.* **34**, 143–150

49. Kim, J. W., Marquez, C. P., Sperberg, R. A. P., Wu, J., Bae, W. G., Huang, P. S., Sweet-Cordero, E. A., and Cochran, J. R. (2020) Engineering a potent receptor superagonist or antagonist from a novel IL-6 family cytokine ligand. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 14110–14118

50. Lui, B. G., Salomon, N., Wustehube-Lausch, J., Daneschdar, M., Schmoldt, H. U., Tureci, O., and Sahin, U. (2020) Targeting the tumor vasculature with engineered cystine-knot miniproteins. *Nat. Commun.* **11**, 295

51. Kimura, R. H., Wang, L., Shen, B., Huo, L., Tummers, W., Filipp, F. V., Guo, H. H., Haywood, T., Abou-Elkacem, L., Baratto, L., Habte, F., Devulapally, R., Witney, T. H., Cheng, Y., Tikole, S., *et al.* (2019) Evaluation of integrin alphavbeta6 cystine knot PET tracers to detect cancer and idiopathic pulmonary fibrosis. *Nat. Commun.* **10**, 4673

52. Getz, J. A., Rice, J. J., and Daugherty, P. S. (2011) Protease-resistant peptide ligands from a knottin scaffold library. *ACS Chem. Biol.* **6**, 837–844

53. Patil, C. G., Walker, D. G., Miller, D. M., Butte, P., Morrison, B., Kittle, D. S., Hansen, S. J., Nufer, K. L., Byrnes-Blake, K. A., Yamada, M., Lin, L. L., Pham, K., Perry, J., Parrish-Novak, J., Ishak, L., *et al.* (2019) Phase 1 safety, pharmacokinetics, and Fluorescence imaging study of Tozuleristide (BLZ-100) in Adults with newly diagnosed or recurrent Gliomas. *Neurosurgery* **85**, E641–E649

54. Tummers, W. S., Kimura, R. H., Abou-Elkacem, L., Beinat, C., Vahrmeijer, A. L., Swijnenburg, R. J., Willmann, J. K., and Gambhir, S. S. (2018) Development and Preclinical Validation of a cysteine knottin peptide targeting integrin alphavbeta6 for Near-infrared Fluorescent-guided Surgery in Pancreatic cancer. *Clin. Cancer Res.* **24**, 1667–1676

55. Grundemann, C., Stenberg, K. G., and Gruber, C. W. (2018) T20K: An immunomodulatory cyclotide on its way to the clinic. *Int. J. Pept. Res. Therap.* **25**, 9–13

56. Craik, D. J., Daly, N. L., Bond, T., and Waine, C. (1999) Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. *J. Mol. Biol.* **294**, 1327–1336

57. Huang, Y. H., Du, Q., and Craik, D. J. (2019) Cyclotides: Disulfide-rich peptide toxins in plants. *Toxicon* **172**, 33–44

58. de Veer, S. J., Kan, M. W., and Craik, D. J. (2019) Cyclotides: From structure to function. *Chem. Rev.* **119**, 12375–12421

59. Gruber, C. W., Elliott, A. G., Ireland, D. C., Delprete, P. G., Dessein, S., Goransson, U., Trabi, M., Wang, C. K., Kinghorn, A. B., Robbrecht, E., and Craik, D. J. (2008) Distribution and evolution of circular miniproteins in flowering plants. *Plant Cell* **20**, 2471–2483

60. Mahatmanto, T., Mylne, J. S., Poth, A. G., Swedberg, J. E., Kaas, Q., Schaefer, H., and Craik, D. J. (2015) The evolution of Momordica cyclic peptides. *Mol. Biol. Evol.* **32**, 392–405

61. Park, S., Yoo, K. O., Marcussen, T., Backlund, A., Jacobsson, E., Rosengren, K. J., Doo, I., and Goransson, U. (2017) Cyclotide evolution: Insights from the analysis of their precursor sequences, structures and distribution in Violets (Viola). *Front Plant Sci.* **8**, 2058

62. Gould, A., and Camarero, J. A. (2017) Cyclotides: Overview and biotechnological applications. *Chembiochem* **18**, 1350–1363

63. Wang, C. K., Kaas, Q., Chiche, L., and Craik, D. J. (2008) CyBase: A database of cyclic protein sequences and structures, with applications in protein discovery and engineering. *Nucleic Acids Res.* **36**, D206–210

64. Saska, I., Gillon, A. D., Hatsugai, N., Dietzgen, R. G., Hara-Nishimura, I., Anderson, M. A., and Craik, D. J. (2007) An asparaginyl endopeptidase mediates *in vivo* protein backbone cyclization. *J. Biol. Chem.* **282**, 29721–29728

65. Nguyen, G. K., Wang, S., Qiu, Y., Hemu, X., Lian, Y., and Tam, J. P. (2014) Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nat. Chem. Biol.* **10**, 732–738

66. Rehm, F. B. H., Jackson, M. A., De Geyter, E., Yap, K., Gilding, E. K., Durek, T., and Craik, D. J. (2019) Papain-like cysteine proteases prepare plant cyclic peptide precursors for cyclization. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 7831–7836

67. Simonsen, S. M., Sando, L., Rosengren, K. J., Wang, C. K., Colgrave, M. L., Daly, N. L., and Craik, D. J. (2008) Alanine scanning mutagenesis of the prototypic cyclotide reveals a cluster of residues essential for bioactivity. *J. Biol. Chem.* **283**, 9805–9813

68. Simonsen, S. M., Daly, N. L., and Craik, D. J. (2004) Capped acyclic permutants of the circular protein kalata B1. *FEBS Lett.* **577**, 399–402

69. Daly, N. L., and Craik, D. J. (2000) Acyclic permutants of naturally occurring cyclic proteins. Characterization of cystine knot and beta-sheet formation in the macrocyclic polypeptide kalata B1. *J. Biol. Chem.* **275**, 19068–19075

70. Atkinson, J. T., Jones, A. M., Nanda, V., and Silberg, J. J. (2019) Protein tolerance to random circular permutation correlates with thermostability and local energetics of residue-residue contacts. *Protein Eng. Des. Sel.* **32**, 489–501

71. Tombling, B. J., Wang, C. K., and Craik, D. J. (2020) EGF-like and other disulfide-rich Microdomains as therapeutic scaffolds. *Angew. Chem. Int. Ed. Engl.* **59**, 11218–11232

72. Wang, C. K., Ghani, H. A., Bundock, A., Weidmann, J., Harvey, P. J., Edwards, I. A., Schroeder, C. I., Swedberg, J. E., and Craik, D. J. (2018) Calcium-mediated Allostery of the EGF fold. *ACS Chem. Biol.* **13**, 1659–1667

73. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains. *Science* **181**, 223–230

74. Morris, E. R., and Searle, M. S. (2012) Overview of protein folding mechanisms: Experimental and theoretical approaches to probing energy landscapes. *Curr. Protoc. Protein Sci.* **Chapter 28**. Unit 28.2.1-22

75. Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (1995) Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* **21**, 167–195

76. Tzul, F. O., Vasilchuk, D., and Makhatadze, G. I. (2017) Evidence for the principle of minimal frustration in the evolution of protein folding landscapes. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E1627–E1632

77. Zhangsun, D., Zhu, X., Kaas, Q., Wu, Y., Craik, D. J., McIntosh, J. M., and Luo, S. (2017) alphaO-Conotoxin GeXIVA disulfide bond isomers exhibit differential sensitivity for various nicotinic acetylcholine receptors but retain potency and selectivity for the human alpha9alpha10 subtype. *Neuropharmacology* **127**, 243–252

78. Grishin, A. A., Wang, C. I., Muttenthaler, M., Alewood, P. F., Lewis, R. J., and Adams, D. J. (2010) Alpha-conotoxin AuIB isomers exhibit distinct inhibitory mechanisms and differential sensitivity to stoichiometry of alpha3beta4 nicotinic acetylcholine receptors. *J. Biol. Chem.* **285**, 22254–22263

79. Dutton, J. L., Bansal, P. S., Hogg, R. C., Adams, D. J., Alewood, P. F., and Craik, D. J. (2002) A new level of conotoxin diversity, a non-native disulfide bond connectivity in alpha-conotoxin AuIB reduces structural definition but increases biological activity. *J. Biol. Chem.* **277**, 48849–48857

80. Porebski, B. T., and Buckle, A. M. (2016) Consensus protein design. *Protein Eng. Des. Sel.* **29**, 245–251

81. Sternke, M., Tripp, K. W., and Barrick, D. (2019) Consensus sequence design as a general strategy to create hyperstable, biologically active proteins. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 11275–11284

82. Risso, V. A., Sanchez-Ruiz, J. M., and Ozkan, S. B. (2018) Biotechnological and protein-engineering implications of ancestral protein resurrection. *Curr. Opin. Struct. Biol.* **51**, 106–115

83. Gumulya, Y., and Gillam, E. M. (2017) Exploring the past and the future of protein evolution with ancestral sequence reconstruction: The 'retro' approach to protein engineering. *Biochem. J.* **474**, 1–19

84. Wheeler, L. C., Lim, S. A., Marqusee, S., and Harms, M. J. (2016) The thermostability and specificity of ancient proteins. *Curr. Opin. Struct. Biol.* **38**, 37–43

85. Buchholz, P. C. F., Zeil, C., and Pleiss, J. (2018) The scale-free nature of protein sequence space. *PLoS One* **13**, e0200815

86. Bastolla, U., Dehouck, Y., and Echave, J. (2017) What evolution tells us about protein physics, and protein physics tells us about evolution. *Curr. Opin. Struct. Biol.* **42**, 59–66

87. Salinas, V. H., and Ranganathan, R. (2018) Coevolution-based inference of amino acid interactions underlying protein function. *Elife* **7**, e34300

88. Sutto, L., Marsili, S., Valencia, A., and Gervasio, F. L. (2015) From residue coevolution to protein conformational ensembles and functional dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13567–13572

89. Kelil, A., Levy, E. D., and Michnick, S. W. (2016) Evolution of domain-peptide interactions to coadapt specificity and affinity to functional diversity. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E3862–3871

90. Zakas, P. M., Brown, H. C., Knight, K., Meeks, S. L., Spencer, H. T., Gaucher, E. A., and Doering, C. B. (2017) Enhancing the pharmaceutical properties of protein drugs by ancestral sequence reconstruction. *Nat. Biotechnol.* **35**, 35–37

91. Wright, C. F., Teichmann, S. A., Clarke, J., and Dobson, C. M. (2005) The importance of sequence diversity in the aggregation and evolution of proteins. *Nature* **438**, 878–881