Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Are natural proteins special? Can we do that?

Michael H Hecht, Shlomo Zarzhitsky, Christina Karas and Sarangan Chari

Abstract

Natural proteins represent a minuscule fraction of possible sequence space. These very rare sequences display remarkable properties: They fold into many different stable structures, and perform a wide range of complex biological functions. These two considerations — rarity and functionality — may suggest that natural proteins are somehow special. Is this true? We address this question by exploring attempts to recapitulate the *special* structures and functions of natural proteins into sequences designed *de novo*.

Address
Departments of Chemistry and Molecular Biology, Princeton University, Princeton, NJ 08540, USA

Corresponding author: Hecht, Michael H (hecht@princeton.edu)

Current Opinion in Structural Biology 2018, **48**:124–132

This review comes from a themed issue on **Proteins**

Edited by **Birte Hocker** and **Jakob Winther**

https://doi.org/10.1016/j.sbi.2017.11.009

Introduction

When we ask whether something is 'special' we are implicitly asking two questions: Is it rare? ('she is one in a million'); and Is it easy/hard to replicate ('I can do *that!*'). Are natural proteins rare? Can we replicate their structures and properties?

For a relatively short protein of 100 amino acids, there are 20^{100} possible sequences. It has been estimated that a collection containing one molecule of all these sequences would fill a volume larger than a mole of universes [1] (Figure 1). While the exact number of existent natural sequences is unknown, it is dwarfed by this number of possible sequences. By this criterion, natural proteins — a miniscule fraction of possible proteins — are rare; far more unusual than your friend who is one in a million.

This unusual collection of natural proteins arose in response to selective pressures. The surviving sequences enhanced the fitness of their hosts, while an almost unimaginable number of alternative sequences were lost to extinction. This may lead one to speculate that the survivors of life-or-death selections that operated over billions of years in myriad cells and organisms must surely be *special*. But are they? *Can we do that?*

Recent advances in genome sequencing, proteomics, protein design, and synthetic biology enable us to address these questions with more data and (hopefully!) more insight than ever before. It is now possible to assess whether we can create entirely novel proteins that recapitulate the key features of naturally evolved proteins. Can we produce non-natural sequences? Will they fold? Will they bind, assemble, and catalyze? Can we create novel proteins that sustain life? Can we do *that?*

Early steps toward protein design

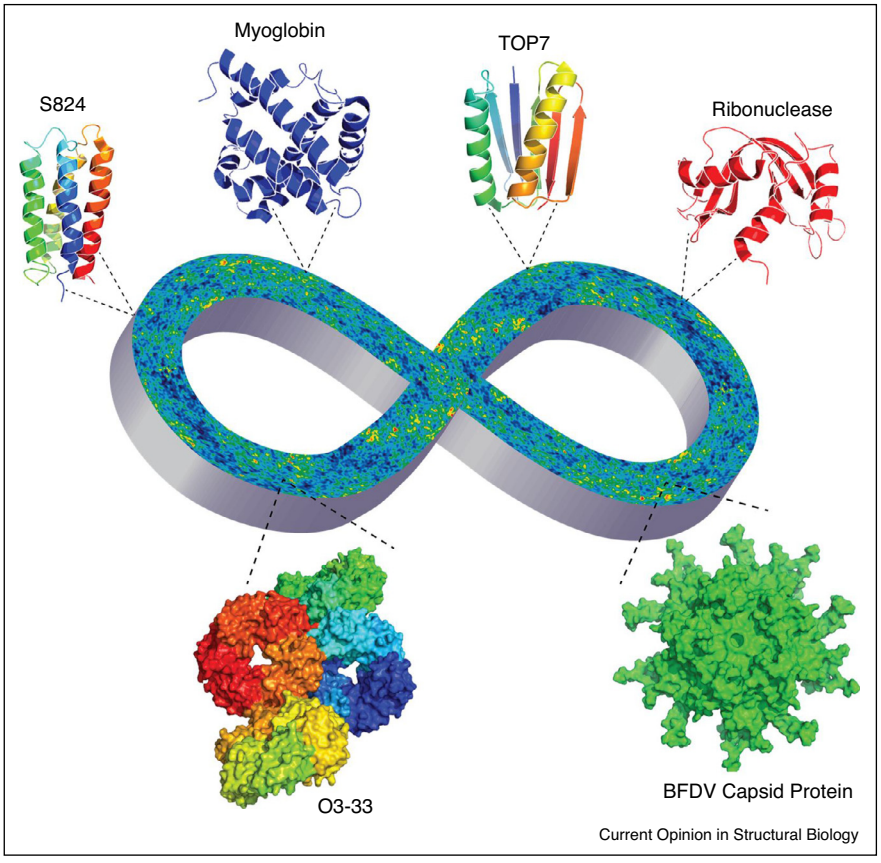
Since the seminal experiments in the 1960s by Anfinsen [2] and Merrifield [3], it has been clear that natural proteins can fold without assistance from any 'life force' provided by living organisms. A quarter of a century later, scientists began to ask whether *non*-natural sequences could also fold spontaneously. The initial goals were rather modest: Can one devise sequences that are unrelated to natural sequences, but nonetheless fold into simple 4-helix bundles? Early successes in the late 1980s and early 1990s included α-4, designed by Regan and DeGrado [4], and Felix, designed by Hecht, Ogden and the Richardsons [5]. These early studies showed that folding *per se* is not a *special* property of natural sequences. Artificial sequences fold too. So at least we can do *that.*

Novel proteins fold into stable structures: both natural and unnatural

While early work on protein design focused on simple structures, the field progressed rapidly, and protein designers soon tackled more challenging problems. Less than a decade after publication of the first 4-helix bundles, Dahiyat and Mayo demonstrated that it is possible to design novel sequences that fold into zinc finger domains, containing alpha structure, beta structure, and a bound metal [6]. Their novel sequence was chosen by a fully automated computational algorithm, and the resulting protein (FSD-1) folded into a structure that closely matched the design target. Importantly, in contrast to the previous designs (α-4 and Felix), FSD-1 formed a well-ordered — rather than dynamic — structure.

While FSD-1 showed that natural *sequences* are not special in their ability to fold into native-like structures, a

Figure 1



A universe of natural and novel proteins. The central image shows the cosmic background radiation of the early universe [13] superimposed on the symbol of infinity. Natural sequences are a miniscule fraction of the astronomical size of possible sequence space. Ribbon diagrams show a natural protein that binds a cofactor (myoglobin, 1MBN); a natural enzyme (ribonuclease, 1FS3): a *de novo* protein from a combinatorial library that folds into a native-like structure (S-824, 1P68); and a computationally designed sequence that folds into a novel structure (TOP7, 1QYS). Space filling models show a natural protein assembly (BFDV capsid protein, 5J37), and a fully designed assembly (03-33, 4DDF).

question remained: Are the natural *structures* that were selected by evolution somehow special? Are other structures and topologies possible? This question was answered by Kuhlman, Baker, and colleagues, when they designed TOP7, a protein with a non-natural sequence that folds into a structure not previously seen in nature [7] (Figure 1).

In the intervening years, a wide range of novel sequences and novel structures have been designed *de novo*. These include idealized α-helical structures, and a range of α/β topologies [8,9•]. Fully β-sheet proteins have also been designed [10]; however, because β-strands are prone to aggregate [11], progress in this area has been slower. Nonetheless, it is clear that both natural and unnatural sequences can fold into a wide range of natural and unnatural structures [12••].

A hallmark of natural proteins is their tendency to fold cooperatively. Although this is not universally true, most natural proteins fold and unfold by a two state mechanism without stable intermediates. In the early days of protein design, this feature seemed special. Designing novel sequences that folded cooperatively was challenging [4,5] and became a gold standard for early workers in the field. However, as the field of protein design matured, many (although not all) novel sequences were shown to fold cooperatively [6–8,14–16]. Thus, it appears that cooperativity *per se* is not a special property of natural proteins. We too can produce cooperative systems.

Novel proteins by the millions and trillions

The preceding sections highlight achievements in the design of individual proteins. Natural proteins, however, were selected from feedstocks containing myriad sequences. Are vast collections of protein sequences a special property of natural ecosystems? Can such collections be generated *de novo?*

A combinatorial library of entirely novel proteins was first described in the early 1990s by Kamtekar *et al.* [17]. A collection of synthetic genes was constructed and cloned into *E. coli* for protein expression. The library was based on a 'binary code' strategy, in which the sequence locations of polar and nonpolar amino acids were specified explicitly, but the precise identities of the side chains were allowed to vary [17]. The combinatorial diversity of the binary pattern was consistent with 5×10^{41} different amino acid sequences. Of course, far fewer genes were actually synthesized, and ultimately, the number of expressed proteins was limited by the efficiency of DNA transformation into *E. coli*, typically $\sim 10^6$. Biophysical studies showed that finding α-helical proteins in these libraries of semi-random sequences was not difficult at all. They were not *special*; indeed *most* of the purified proteins formed α-helical bundles. Subsequent libraries of binary patterned sequences showed that highly stable and well-ordered structures could readily be found in collections of non-natural proteins [14–16] (Figure 1). Thus, folding into natural-like structures does not require natural-like sequences. Moreover — at least in some cases — folding into natural-like structures does not require unnatural sequences to be designed atom-by-atom by rational or computational methods.

While binary patterned libraries showed that large collections of unevolved sequences can fold (and function — *see below*), isolation of these proteins still depends on expression in natural bacterial cells. This limitation was circumvented in seminal work by Keefe and Szostak, who showed that enormous libraries of novel sequences can be generated without cells [18]. By using *in vitro* RNA display technology developed in the Szostak laboratory [19], they produced 6×10^{12} novel 80-residue sequences. Next they selected — also in a cell free system — individual proteins that bound ATP. More recent advances show that even larger libraries can be generated and subjected to selections without any requirement for living cells [20].

Assembly into nanostructures and molecular machines

Natural proteins do not merely fold *intra*molecularly; they also assemble *inter*molecularly. Assembly into oligomers and molecular machines requires proteins to recognize sites on other proteins (or other copies of the same protein). The recognition sites on surfaces of natural proteins were honed by eons of evolutionary selection. Is natural selection necessary to achieve the finely tuned affinities and specificities required for correct assembly?

This question was probed by DeLano *et al.*, who compared binding by natural and *de novo* sequences to a patch on the hinge region of the Fc fragment of immunoglobulin G [21]. At least four different natural protein structures recognize this patch as the preferred binding site on the Fc fragment. Is the ability to recognize this patch a *special* feature of the natural proteins that co-evolved with it? Or would sequences that did not arise in nature also prefer this patch relative to other loci on the Fc fragment? DeLano *et al.* answered this question using phage display to probe collections of novel sequences for their ability to bind the Fc fragment. Biophysical studies and crystal structures showed that the *de novo* sequence that bound the Fc fragment recognized the exact same patch; moreover, it bound using interactions that mimicked those used by the natural binding partners. Thus, not only is protein assembly *possible* for non-natural sequences, but even the specific binding used by natural proteins can be recapitulated by non-natural sequences.

The experiment described above probed the ability of non-natural sequences to recognize a binding patch on a natural protein. More recent studies have created assemblies in which the entire recognition process (on *both* sides) is mediated by sequences that did not benefit from natural selection. For example, Kobayashi *et al.* created self-assembling nano-architectures using domain-swapped dimers of novel 4-helix bundles [22]. The intermolecular interactions at the dimer interface were mediated entirely by helix/helix contacts contributed by non-natural sequences.

In some designed systems, the interactions mediated by *de novo* sequences look familiar — *i.e.* they resemble the types of protein/protein interactions seen in natural assemblies. In others, however, the contacts designed into the non-natural sequences are very different from those found in nature. For example, Tezcan and coworkers harnessed the specific and well-understood coordination geometry of transition metals to construct numerous metal-directed protein assemblies that differ significantly from naturally occurring assemblies [23–26]. In another example of non-natural recognition, Boyken *et al.* designed homo-oligomers of α-helices in which the inter-helical contacts are formed by arrays of hydrogen-bonded networks [27•]. This contrasts with most natural protein/protein interfaces, which are typically mediated by hydrophobic packing with some nearby polar interactions. Boyken *et al.* note that 'even with the tremendous diversity observed in nature, there are fundamentally new modes of interaction to be discovered in proteins' [27•]. Perhaps, it is the *non*-natural proteins that are special!

The possibility of systematically building very large and sophisticated protein assemblies using non-natural binding interfaces has been brought to fruition by several recent studies. Woolfson and coworkers designed interfaces into a series of novel α-helices, ultimately leading to self-assembling barrels and cages [28,29•]. More recently, collaborative work from the Yeates and Baker groups reported the computational design of icosahedral assemblies with molecular weights exceeding 10^6 Daltons and

diameters up to 40 nanometers [30•,31,32••]. The geometries and dimensions of these assemblies resemble those of viral capsids, suggesting that even structures selected by evolution to carry hereditary information between organisms can be devised *de novo* in the laboratory (Figure 1).

Is binding special?

Most natural proteins bind small molecules and/or metals [33,34]. Indeed, without binding, catalysis would not be possible. Are non-natural sequences also capable of binding?

Skolnick *et al.* addressed this question using a computational approach [35]. They generated an *in silico* library of compact artificial proteins with lengths ranging from 40 to 250 residues. Next, they asked whether binding sites for small molecules, proteins, or DNA could be found in this collection of hypothetical proteins. Remarkably, they found many native-like binding sites in this collection of sequences designed *in silico* solely for structure, but not for function. Based on these findings, they suggested the biochemical function of natural proteins may not be special, and may be 'an intrinsic feature of proteins which nature has significantly optimized during evolution' [35].

Experimental attempts to install binding sites into novel proteins initially focused on transition metals and heme cofactors, and pioneering studies showed that rational and computational methods could yield novel proteins that bind metals and/or heme [36–40]. It turns out that binding metals or heme is *not* difficult, and >25% of unselected binary patterned proteins (*see above*) are capable of binding [39,41]. Although these sequences from semi-random libraries do not bind with high affinity or specificity, they suggest that some level of binding in ancestral natural proteins may have been rather common.

More recent studies used rational and computational methods to design tight and specific binding to metals and/or small molecules. For example, as noted above, Tezcan and coworkers engineered proteins to bind metals at precise and specific loci [23–26]. In addition, Dutton and coworkers designed a series of α-helical protein maquettes that bind a range of biologically significant metals and cofactors [42–44]. With the goal of achieving very high affinity and specificity, Tinberg *et al.* used Rosetta to redesign natural protein scaffolds to bind digoxigenin with preordered shape complementarity [45•]. Most recently, DeGrado and coworkers used a mathematically parameterized *de novo* backbone rather than a natural protein scaffold to design a protein ligand complex with sub-Angstrom accuracy [46••]. The ligand chosen for their design was a non-natural porphyrin, thereby demonstrating the possibility of constructing novel holo-proteins in which the sequence, structure, and cofactor are *all* artificial and not natural. These and other studies show that specific binding to small molecules is not a *special* feature that arises only in response to eons of natural selection; rational and computational methods *can do that* too.

These results and related studies led Cherny *et al.* to ask, "Is the ability to bind small molecules a property that arises only in response to biological selection or computational design? Or . . . is small molecule binding a property of folded proteins that occurs readily amidst collections of unevolved sequences?" [47]. The availability of large collections of *de novo* proteins presented a new opportunity to answer this question. Cherny *et al.* used small molecule microarrays to test whether proteins from a combinatorial library of novel sequences would bind any of 10 000 different molecules displayed on the array. They found that several of the *de novo* proteins bound several of the small molecules with moderate affinities and specificities. This is particularly surprising since the binary patterned sequences had been designed only for structure, and not for function. Apparently, the ability to bind small molecules is *not* a special feature, and requires neither millennia of evolution, nor days of CPU time.

In addition to binding small molecules and metals, natural proteins bind macromolecules with high levels of affinity and discrimination. The remarkable ability of the mammalian immune system to generate antibodies in response to virtually any macromolecular challenge seems rather special. However, antibodies are also 'special' in a negative way: They are large multi-chained proteins that are difficult to express. Moreover once expressed, they are not stable to extreme conditions, and have short shelf lives. These properties challenged protein designers to devise novel sequences and structures that possess the favorable properties (affinity and specificity), but not the unfavorable properties (large size, difficult expression and low stability) of natural antibodies. Using Rosetta computational design, Baker and coworkers have made dramatic progress toward this goal. In an initial foray into this area, Fleischmann *et al.* designed "disembodied amino acid residues" to form binding patches that recognized a conserved stem region on influenza hemagglutinin [48]. Then they searched the database of natural protein structures to find scaffolds upon which these disembodied residues could be grafted. This approach led to novel proteins (on natural scaffolds) that bound the target with low nanomolar affinities using structural interfaces nearly identical to those in the designed model [48]. More recently, Chevalier *et al.* brought the design of binding interfaces to a new level. By building on recent developments in gene synthesis, computational design, high throughput screening, and next-generation sequencing, they synthesized >22 000 *de novo* mini-proteins designed to fold into a range of different topologies. Next, they tested the novel sequences for binding to

hemagglutinin or botulinum toxin [49••]. Ultimately, they isolated many novel sequences and structures that bound these targets with high affinity. Moreover, the novel mini-proteins protected mice from influenza virus. These results, which would have been unimaginable just a few years ago, suggest that even the remarkable binding capabilities of the mammalian immune system may not be so special after all.

Are natural enzymes special?

Catalysis is one of the core concerns of modern chemistry, and synthetic chemists take great pride in discovering potent catalysts for important reactions. The best catalysts, however, were not designed by chemists; they were evolved by nature. Indeed, some natural enzymes are so effective, they have been described as 'perfect' catalysts [50]. Is this level of catalytic proficiency special? Can biochemists devise non-natural enzymes capable of catalyzing difficult chemical transformations?

Several approaches have been used to construct novel enzymes. A simple and appealing strategy is to engineer binding sites for metals and/or cofactors into the structures of natural protein scaffolds. Cofactors often possess some level of intrinsic activity, and sequestration into a protein can both enhance activity, and impart substrate specificity. Nature pioneered this strategy, and a relatively small number of metals and cofactors have been used as 'pre-organized activity modules' [51] to generate an enormous variety of natural enzymes. By following nature's lead, chemists and biochemists have developed new generations of biocatalysts, in which cofactors are engineered into natural proteins to produce catalytic activities that differ substantially from those in nature [52•,53••,54,55,56•,57,58].

A more challenging goal is to use metals and cofactors to impart activity into novel sequences that did *not* arise in nature. By designing cofactor-binding sites into their α-helical maquettes (*see above*), the Dutton group succeeded in producing novel enzymes that catalyze a range of oxidoreductase activities [42,43,59•]. Other groups have also bound metals and cofactors into novel sequences to generate novel cofactor-dependent enzymatic activities [51,60–62].

Perhaps the most challenging goal in creating non-natural enzymes is to design a novel active site without using cofactors. This is considerably more difficult than designing the structures and assemblies summarized in the preceding sections. While designing structure requires precision at Angstrom resolution, producing efficient catalysts may require precision at fractions of an Angstrom [63••]. Pioneering studies aimed at this goal were reported over the past decade [64–66]. At first glance these successes seemed spectacular. However, closer inspection revealed they had rather modest turnovers, nowhere near the special catalytic rates enabled by naturally evolved enzymes [67,68].

Recently, however, several groups, most notably Hilvert and colleagues, subjected these computationally designed enzymes to laboratory-based evolution. After many rounds of mutagenesis and selection, they evolved some of these novel catalysts to the point where they rival the activities of natural enzymes [63••,65,66,68,69••]. These results led Obexer *et al.* to suggest "there is nothing magical about the catalytic activities or mechanisms of naturally occurring enzymes, or the evolutionary process that gave rise to them" [69••]. Not magical. Perhaps not even special.

It is difficult to create novel enzymes with high levels of activity. Yet it is important to realize that nature did not start with perfect enzymes. Ancestral enzymes, presumably, had very low levels of activity [70]. This leads us to ask whether weakly catalytic proteins — as compared to proficient enzymes — occur in collections of novel sequences that have neither been subjected to evolution nor to computational design. Several studies addressed this question, both *in vitro* and *in vivo*. For example, Haehnel and coworkers used solid phase synthesis to produce hundreds of 4-helix bundles affixed to a solid template [71]. Screening these libraries for various catalytic activities led to the discovery of low levels of heme oxygenase activity [71]. An alternative approach using expression *in vivo* found that binary patterned 4-helix bundles (*see above*) had low levels of esterase and lipase activity. Moreover, in the presence of heme, a number of them have moderate levels of peroxidase activity [41,51]. However, further design and/or laboratory-based evolution [72] will be required before any of these protein catalysts approach the special proficiency of enzymes.

Special proteins sustain life

Evolution does not explicitly select for folding, assembly, binding, or catalysis. Evolution selects for survival; and the properties listed above arose collaterally as a consequence of nature's selection for sequences that sustain life. Therefore, when we ask "Are Natural Proteins Special? *Can We Do That?*" perhaps we should ask whether we can construct non-natural proteins that sustain the growth of living cells.

While this sounds like a tall order, experiments addressing this question are straightforward. One simply needs to knock out a conditionally essential gene (a gene essential for life under a certain set of conditions) in a simple model organism such as *E. coli*, and ask whether addition of sequences from a library of genes encoding novel proteins can rescue the deletion, and sustain cell growth under those conditions. Our lab has performed this experiment using several knockouts in *E. coli*, and a library of a million binary patterned proteins. In four

cases, we found novel proteins that rescued these knockouts. The natural proteins deleted in the knockout strains had different activities, including phosphoserine phosphatase, citrate synthase, threonine deaminase, and enterobactin esterase [73]. Rescuing deletions of these enzymes suggests that a novel protein is enzymatically active *in vivo*. However, this is not necessarily the case; it is also possible to rescue deletions by altering the regulation or activity of endogenous proteins. Further experiments showed that in some cases the novel proteins rescue by altering the regulation of natural *E. coli* proteins with promiscuous activities [74••,75•]. In other cases, a novel protein sustains life by functioning *in vivo* as a *de novo* enzyme. It is not as active as the natural protein, but it is *special* enough to sustain life [76••].

Evolution of specialists

This volume is entitled "Proteins: An Evolutionary Perspective", and the current chapter asks "Are Natural Proteins Special?" In this last section, we combine these concepts and ask whether the evolutionary processes that occur in nature are themselves special. While the previous sections argued that the structures and properties of natural proteins may not be special, and can be recapitulated by non-natural sequences, here we ask whether the very process that gave rise to natural proteins — Darwinian evolution — is special. Can this process be recreated in the laboratory and applied to *de novo* sequences that did not arise in nature?

Many studies have subjected natural sequences to laboratory evolution with the goal of selecting altered properties [57,58,77–79]. Those projects start with sequences that themselves are products of evolutionary history in the wild. Hence, they are biased by remnants of evolutionary baggage, which arose over eons of natural selection in organisms that evolved in environments that may not be relevant to the traits under selection in the laboratory.

It would be interesting to apply evolutionary strategies to sequences created entirely *de novo*, which by definition, are not encumbered by historical baggage. A particularly compelling experiment would be to test a fundamental premise in evolutionary theory pertaining to the evolution of *specialists* from generalists. Such a premise was put forth by Jensen, who proposed that although ancestral proteins were not very active, they "possessed a very broad specificity, permitting them to react with a wide range of related substrates." According to Jensen, this was important early in evolution to "maximize the catalytic versatility of an ancestral cell that functioned with limited enzyme resources" [70]. As evolution selected for greater levels of activity, proteins would have evolved from poorly active generalists to highly active specialists. Is this true? It is difficult to test this hypothesis with natural proteins because their ancestral sequences are lost to time. However, we *can* test these ideas with novel proteins. In one example, a binary patterned sequence that rescued two different auxotrophs of *E. coli* was evolved in the laboratory — in two separate experiments — for better rescue of each auxotroph independently. Sure enough, as activity increased, promiscuity decreased [80•]. Thus, even evolutionary processes performed on non-natural proteins in non-natural laboratory settings can lead a promiscuous activity that is general to evolve into a non-promiscuous activity that is *special*.

In summary, the results reviewed here suggest that natural proteins may not be so special after all. We *can* replicate their structures and properties using both *de novo* design, and "An Evolutionary Perspective."

Conflicts of interest

The authors have nothing to disclose.

Acknowledgements

This work was supported in part by NSF Grant MCB-1409402. The authors thank Rayleen Hu and Clara Yu for their enthusiastic participation.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Beasley JR, Hecht MH: **Protein design: the choice of de novo sequences**. *J Biol Chem* 1997, **272**:2031-2034 http://dx.doi.org/10.1074/jbc.272.4.2031.

2. Haber E, Anfinsen CB: **Side-chain interactions governing the pairing of half-cystine residues in ribonuclease**. *J Biol Chem* 1962, **237**:1839-1844.

3. Gutte B, Merrifield R: **Total synthesis of an enzyme with ribonuclease a activity**. *J Am Chem Soc* 1969, **91**:501-502 http://dx.doi.org/10.1021/ja01030a050.

4. Regan L, Degrado W: **Characterization of a helical protein designed from 1st principles**. *Science* 1988, **241**:976-978 http://dx.doi.org/10.1126/science.3043666.

5. Hecht M, Richardson J, Richardson D, Ogden R: **De novo design, expression, and characterization of felix: a four-helix bundle protein of native-like sequence**. *Science* 1990, **249**:884-891 http://dx.doi.org/10.1126/science.2392678.

6. Dahiyat B, Mayo S: **De novo protein design: fully automated sequence selection**. *Science* 1997, **278**:82-87 http://dx.doi.org/10.1126/science.278.5335.82.

7. Kuhlman B, Dantas G, Ireton G, Varani G, Stoddard B, Baker D: **Design of a novel globular protein fold with atomic-level accuracy**. *Science* 2003, **302**:1364-1368 http://dx.doi.org/10.1126/science.1089427.

8. Koga N, Tatsumi-Koga R, Liu G, Xiao R, Acton TB, Montelione GT, Baker D: **Principles for designing ideal protein structures**. *Nature* 2012, **491**:222-227 http://dx.doi.org/10.1038/nature11600.

9. • Lin Y, Koga N, Tatsumi-Koga R, Liu G, Clouser AF, Montelione GT, Baker D: **Control over overall shape and size in de novo designed proteins**. *Proc Natl Acad Sci U S A* 2015, **112**:E5478-E5485 http://dx.doi.org/10.1073/pnas.1509508112.
Design principles focused on consecutive regions of secondary structure allow the modulation of protein size and shape within a given fold.

10. West M, Wang W, Patterson J, Mancias J, Beasley J, Hecht M: **De novo amyloid proteins from designed combinatorial libraries**. *Proc Natl Acad Sci U S A* 1999, **96**:11211-11216 http://dx.doi.org/10.1073/pnas.96.20.11211.

11. Wang W, Hecht M: **Rationally designed mutations convert de novo amyloid-like fibrils into monomeric beta-sheet proteins**. *Proc Natl Acad Sci U S A* 2002, **99**:2760-2765 http://dx.doi.org/10.1073/pnas.052706199.

12. •• Rocklin GJ, Chidyausiku TM, Goreshnik I, Ford A, Houliston S, Lemak A, Carter L, Ravichandran R, Mulligan VK, Chevalier A *et al.*: **Global analysis of protein folding using massively parallel design, synthesis, and testing**. *Science* 2017, **357**:168-174 http://dx.doi.org/10.1126/science.aan0693.
High-throughput protein design, gene synthesis, expression, and protein stability assays enable the production of thousands of *de novo* sequences for four different folds. Large collections enable examination of sequence-structure relationship in previously unexplored sequence space.

13. Bennett C, Halpern M, Hinshaw G, Jarosik N, Kogut A, Limon M, Meyer S, Page L, Spergel D, Tucker G *et al.*: **First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: preliminary maps and basic results**. *Astrophys J Suppl Ser* 2003, **148**:1-27 http://dx.doi.org/10.1086/377253.

14. Wei Y, Kim S, Fela D, Baum J, Hecht M: **Solution structure of a de novo protein from a designed combinatorial library**. *Proc Natl Acad Sci U S A* 2003, **100**:13270-13273 http://dx.doi.org/10.1073/pnas.1835644100.

15. Wei Y, Liu T, Sazinsky S, Moffet D, Pelczer I, Hecht M: **Stably folded de novo proteins from a designed combinatorial library**. *Protein Sci* 2003, **12**:92-102 http://dx.doi.org/10.1110/ps.0228003.

16. Go A, Kim S, Baum J, Hecht MH: **Structure and dynamics of de novo proteins from a designed superfamily of 4-helix bundles**. *Protein Sci* 2008, **17**:821-832 http://dx.doi.org/10.1110/ps.073377908.

17. Kamtekar S, Schiffer J, Xiong H, Babik J, Hecht M: **Protein design by binary patterning of polar and nonpolar amino-acids**. *Science* 1993, **262**:1680-1685 http://dx.doi.org/10.1126/science.8259512.

18. Keefe A, Szostak J: **Functional proteins from a random-sequence library**. *Nature* 2001, **410**:715-718 http://dx.doi.org/10.1038/35070613.

19. Roberts R, Szostak J: **RNA-peptide fusions for the in vitro selection of peptides and proteins**. *Proc Natl Acad Sci U S A* 1997, **94**:12297-12302 http://dx.doi.org/10.1073/pnas.94.23.12297.

20. Golynskiy MV, Haugner JCI, Seelig B: **Highly diverse protein library based on the ubiquitous $(\beta/\alpha)_8$ enzyme fold yields well-structured proteins through in vitro folding selection**. *Chembiochem* 2013, **14**:1553-1563 http://dx.doi.org/10.1002/cbic.201300326.

21. DeLano W, Ultsch M, de Vos A, Wells J: **Convergent solutions to binding at a protein–protein interface**. *Science* 2000, **287**:1279-1283 http://dx.doi.org/10.1126/science.287.5456.1279.

22. Kobayashi N, Yanase K, Sato T, Unzai S, Hecht MH, Arai R: **Self-assembling nano-architectures created from a protein nano-building block using an intermolecularly folded dimeric de novo protein**. *J Am Chem Soc* 2015, **137**:11285-11293 http://dx.doi.org/10.1021/jacs.5b03593.

23. Salgado EN, Faraone-Mennella J, Tezcan FA: **Controlling protein–protein interactions through metal coordination: assembly of a 16-helix bundle protein**. *J Am Chem Soc* 2007, **129**:13374-13375 http://dx.doi.org/10.1021/ja075261o.

24. Salgado EN, Lewis RA, Faraone-Mennella J, Tezcan FA: **Metal-mediated self-assembly of protein superstructures: influence of secondary interactions on protein oligomerization and aggregation**. *J Am Chem Soc* 2008, **130**:6082-6084 http://dx.doi.org/10.1021/ja8012177.

25. Radford RJ, Tezcan FA: **A superprotein triangle driven by nickel (II) coordination: exploiting non-natural metal ligands in protein self-assembly**. *J Am Chem Soc* 2009, **131**:9136-9137 http://dx.doi.org/10.1021/ja9000695.

26. Brodin JD, Medina-Morales A, Ni T, Salgado EN, Ambroggio XI, Tezcan FA: **Evolution of metal selectivity in templated protein interfaces**. *J Am Chem Soc* 2010, **132**:8610-8617 http://dx.doi.org/10.1021/ja910844n.

27. • Boyken SE, Chen Z, Groves B, Langan RA, Oberdorfer G, Ford A, Gilmore JM, Xu C, DiMaio F, Pereira JH *et al.*: **De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity**. *Science* 2016, **352**:680-687 http://dx.doi.org/10.1126/science.aad8865.
The design of specific hydrogen-bond networks enables the creation of dimers, trimers, and tetramers.

28. Fletcher JM, Harniman RL, Barnes FRH, Boyle AL, Collins A, Mantell J, Sharp TH, Antognozzi M, Booth PJ, Linden N *et al.*: **Self-assembling cages from coiled-coil peptide modules**. *Science* 2013, **340**:595-599 http://dx.doi.org/10.1126/science.1233936.

29. • Thomson AR, Wood CW, Burton AJ, Bartlett GJ, Sessions RB, Brady RL, Woolfson DN: **Computational design of water-soluble alpha-helical barrels**. *Science* 2014, **346**:485-488 http://dx.doi.org/10.1126/science.1257452.
Geometrical and computational methods for designing alpha-helical barrels utilizing coiled-coil packing.

30. • King NP, Bale JB, Sheffler W, McNamara DE, Gonen S, Gonen T, Yeates TO, Baker D: **Accurate design of co-assembling multi-component protein nanomaterials**. *Nature* 2014, **510**:103-108 http://dx.doi.org/10.1038/nature13404.
Computational design method for creating nanostructures composed of multiple protein subunits.

31. King NP, Sheffler W, Sawaya MR, Vollmar BS, Sumida JP, Andre I, Gonen T, Yeates TO, Baker D: **Computational design of self-assembling protein nanomaterials with atomic level accuracy**. *Science* 2012, **336**:1171-1174 http://dx.doi.org/10.1126/science.1219364.

32. •• Bale JB, Gonen S, Liu Y, Sheffler W, Ellis D, Thomas C, Cascio D, Yeates TO, Gonen T, King NP, Baker D: **Accurate design of megadalton-scale two-component icosahedral protein complexes**. *Science* 2016, **353**:389-394 http://dx.doi.org/10.1126/science.aaf8818.
Computational design and characterization of protein nanostructures with capsid-like properties — up to 40 nm diameter, rapid self-assembly, and ability to encapsulate cargo.

33. Thomson AJ, Gray HB: **Bio-inorganic chemistry**. *Curr Opin Chem Biol* 1998, **2**:155-158 http://dx.doi.org/10.1016/S1367-5931(98)80056-2.

34. Gray H: **Biological inorganic chemistry at the beginning of the 21st century**. *Proc Natl Acad Sci U S A* 2003, **100**:3563-3568 http://dx.doi.org/10.1073/pnas.0730378100.

35. Skolnick J, Gao M, Zhou H: **How special is the biochemical function of native proteins?** *F1000Res* 2016, **5** http://dx.doi.org/10.12688/f1000research.7374.1.

36. Petrik ID, Liu J, Lu Y: **Metalloenzyme design and engineering through strategic modifications of native protein scaffolds**. *Curr Opin Chem Biol* 2014, **19**:67-75 http://dx.doi.org/10.1016/j.cpba.2014.01.006.

37. Calhoun JR, Kono H, Lahr S, Wang W, DeGrado WF, Saven JG: **Computational design and characterization of a monomeric helical dinuclear metalloprotein**. *J Mol Biol* 2003, **334**:1101-1115.

38. Ghirlanda G, Osyczka A, Liu W, Antolovich M, Smith K, Dutton P, Wand A, DeGrado W: **De novo design of a D-2-symmetrical protein that reproduces the diheme four-helix bundle in cytochrome bc(1)**. *J Am Chem Soc* 2004, **126**:8141-8147 http://dx.doi.org/10.1021/ja039935g.

39. Rojas NRL, Kamtekar S, Simons CT, Mclean JE, Vogel KM, Spiro TG, Farid RS, Hecht MH: **De novo heme proteins from designed combinatorial libraries**. *Protein Sci* 1997, **6**:2512-2524 http://dx.doi.org/10.1002/pro.5560061204.

40. Gibney BR, Dutton PL: **De novo design and synthesis of heme proteins**. *Adv Inorg Chem* 2000, **51**:409-456.

41. Patel SC, Bradley LH, Jinadasa SP, Hecht MH: **Cofactor binding and enzymatic activity in an unevolved superfamily of de novo designed 4-helix bundle proteins**. *Protein Sci* 2009, **18**:1388-1400 http://dx.doi.org/10.1002/pro.147.

42. Robertson DE, Farid RS, Moser CC, Urbauer JL, Mulholland SE, Pidikiti R, Lear JD, Wand AJ, DeGrado WF, Dutton PL: **Design and synthesis of multi-haem proteins**. *Nature* 1994, **368**:425.

43. Fry BA, Solomon LA, Leslie Dutton P, Moser CC: **Design and engineering of a man-made diffusive electron-transport protein**. *Biochim Biophys Acta* 2016, **1857**:513-521.

44. Kodali G, Mancini JA, Solomon LA, Episova TV, Roach N, Hobbs CJ, Wagner P, Mass OA, Aravindu K, Barnsley JE *et al.*: **Design and engineering of water-soluble light-harvesting protein maquettes**. *Chem Sci* 2017, **8**:316-324 http://dx.doi.org/10.1039/C6SC02417C.

45. • Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, Jankowski W, Kalodimos CG, Johnsson K, Stoddard BL, Baker D: **Computational design of ligand-binding proteins with high affinity and selectivity**. *Nature* 2013, **501**:212-216 http://dx.doi.org/10.1038/nature12443.

Computational design method for creating ligand–protein interactions.

46. •• Polizzi NF, Wu Y, Lemmin T, Maxwell AM, Zhang S, Rawson J, Beratan DN, Therien MJ, DeGrado WF: **De novo design of a hyperstable non-natural protein–ligand complex with sub-Å accuracy**. *Nat Chem* 2017 http://dx.doi.org/10.1038/nchem.2846. [advance online publication].

Computational design of an entirely novel protein that binds a non-natural cofactor. The complex is extremely stable and its structure was determined at high resolution.

47. Cherny I, Korolev M, Koehler AN, Hecht MH: **Proteins from an unevolved library of de novo designed sequences bind a range of small molecules**. *ACS Synth Biol* 2012, **1**:130-138 http://dx.doi.org/10.1021/sb200018e.

48. Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch E, Wilson IA, Baker D: **Computational design of proteins targeting the conserved stem region of influenza hemagglutinin**. *Science* 2011, **332**:816.

49. •• Chevalier A, Silva D, Rocklin GJ, Hicks DR, Vergara R, Murapa P, Bernard SM, Zhang L, Lam K, Yao G *et al.*: **Massively parallel de novo protein design for targeted therapeutics**. *Nature* 2017, **550**:74.

Describes the design, creation, and testing of thousands of small *de novo* proteins for binding to clinically relevant targets. The designed miniproteins protected mice from influenza.

50. Knowles J, Albery W: **Perfection in enzyme catalysis — energetics of triosephosphate isomerase**. *Acc Chem Res* 1977, **10**:105-111 http://dx.doi.org/10.1021/ar50112a001.

51. Moffet D, Certain L, Smith A, Kessel A, Beckwith K, Hecht M: **Peroxidase activity in heme proteins derived from a designed combinatorial library**. *J Am Chem Soc* 2000, **122**:7612-7613 http://dx.doi.org/10.1021/ja001198q.

52. • Key HM, Dydio P, Clark DS, Hartwig JF: **Abiological catalysis by artificial haem proteins containing noble metals in place of iron**. *Nature* 2016, **534**:534-537 http://dx.doi.org/10.1038/nature17968.

Replacement of iron with abiological metals in porphyrin proteins enables the catalysis of reactions distinct from those of natural metalloenzymes.

53. •• Dydio P, Key HM, Nazarenko A, Rha JY-, Seyedkazemi V, Clark DS, Hartwig JF: **An artificial metalloenzyme with the kinetics of native enzymes**. *Science* 2016, **354**:102-106 http://dx.doi.org/10.1126/science.aah4427.

Directed evolution used to improve the poor activity of an artificial metalloenzyme, resulting in parameters similar to natural enzymes.

54. Ringenberg MR, Ward TR: **Merging the best of two worlds: artificial metalloenzymes for enantioselective catalysis**. *Chem Commun* 2011, **47**:8470-8476 http://dx.doi.org/10.1039/c1cc11592h.

55. Hyster TK, Knoerr L, Ward TR, Rovis T: **Biotinylated Rh(III) complexes in engineered streptavidin for accelerated asymmetric C-H activation**. *Science* 2012, **338**:500-503 http://dx.doi.org/10.1126/science.1226132.

56. • Emmanuel MA, Greenberg NR, Oblinsky DG, Hyster TK: **Accessing non-natural reactivity by irradiating nicotinamide-dependent enzymes with light**. *Nature* 2016, **540**:414-417 http://dx.doi.org/10.1038/nature20569.

Incorporation of photoexcitable cofactors allows for catalysis of novel reactions.

57. Coelho PS, Brustad EM, Kannan A, Arnold FH: **Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes**. *Science* 2013, **339**:307-310 http://dx.doi.org/10.1126/science.1231434.

58. Kan SBJ, Lewis RD, Chen K, Arnold FH: **Directed evolution of cytochrome c for carbon-silicon bond formation: bringing silicon to life**. *Science* 2016, **354**:1048-1051 http://dx.doi.org/10.1126/science.aah6219.

59. • Dutton LP, Kodali G, Mancini JA, Ennist NM, Stayrook S, Zhao Z, Englander M, Sheehan MM, Fry BA, Bialas C *et al.*: **Toward the biogenesis of manmade oxidoreductases working in cells**. *Biochim Biophys Acta Bioenerg* 2014, **1837**:e9-e10.

Describes progress toward identifying *de novo* oxidoreductases that can be incorporated into living systems. Alpha-helical maquettes can catalyze a range of natural functions and can be expressed in *E. coli* with the appropriate cofactor integration.

60. Nanda V, Rosenblatt M, Osyczka A, Kono H, Getahun Z, Dutton P, Saven J, DeGrado W: **De novo design of a redox-active minimal rubredoxin mimic**. *J Am Chem Soc* 2005, **127**:5804-5805 http://dx.doi.org/10.1021/ja050553f.

61. Mills JH, Khare SD, Bolduc JM, Forouhar F, Mulligan VK, Lew S, Seetharaman J, Tong L, Stoddard BL, Baker D: **Computational design of an unnatural amino acid dependent metalloprotein with atomic level accuracy**. *J Am Chem Soc* 2013, **135**:13393-13399 http://dx.doi.org/10.1021/ja403503m.

62. Der BS, Edwards DR, Kuhlman B: **Catalysis by a de novo zinc-mediated protein interface: implications for natural enzyme evolution and rational enzyme engineering**. *Biochemistry (NY)* 2012, **51**:3933-3940 http://dx.doi.org/10.1021/bi201881p.

63. •• Blomberg R, Kries H, Pinkas DM, Mittl PRE, Gruetter MG, Privett HK, Mayo SL, Hilvert D: **Precision is essential for efficient catalysis in an evolved Kemp eliminase**. *Nature* 2013, **503**:418-421 http://dx.doi.org/10.1038/nature12623.

Computational design and directed evolution of a Kemp eliminase shows it is feasible to emulate the high rate enhancements of natural enzymes.

64. Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O *et al.*: **Kemp elimination catalysts by computational enzyme design**. *Nature* 2008, **453**:190-195 http://dx.doi.org/10.1038/nature06879.

65. Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, Clair JLS, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL *et al.*: **Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction**. *Science* 2010, **329**:309-313 http://dx.doi.org/10.1126/science.1190239.

66. Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas, Carlos F III *et al.*: **De novo computational design of retro-aldol enzymes**. *Science* 2008, **319**:1387-1391 http://dx.doi.org/10.1126/science.1152692.

67. Korendovych IV, DeGrado WF: **Catalytic efficiency of designed catalytic proteins**. *Curr Opin Struct Biol* 2014, **27**:113-121.

68. Kries H, Blomberg R, Hilvert D: **De novo enzymes by computational design**. *Curr Opin Chem Biol* 2013, **17**:221-228.

69. •• Obexer R, Godina A, Garrabou X, Mittl PRE, Baker D, Griffiths AD, Hilvert D: **Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase**. *Nat Chem* 2017, **9**:50-56 http://dx.doi.org/10.1038/NCHEM.2596.

Directed evolution via a droplet-based microfluidic screen improved the activity of an artificial aldolase to rival that of natural enzymes. The active site was extensively modified and a catalytic tetrad arose, demonstrating that precise catalytic centers are not unique to natural proteins.

70. Jensen R: **Enzyme recruitment in evolution of new function**. *Annu Rev Microbiol* 1976, **30**:409-425 http://dx.doi.org/10.1146/annurev.mi.30.100176.002205.

71. Monien BH, Drepper F, Sommerhalter M, Lubitz W, Haehnel W: **Detection of heme oxygenase activity in a library of four-helix bundle proteins: towards the de novo synthesis of functional heme proteins**. *J Mol Biol* 2007, **371**:739-753.

72. Patel SC, Hecht MH: **Directed evolution of the peroxidase activity of a de novo-designed protein**. *Protein Eng Des Sel* 2012, **25**:445-451 http://dx.doi.org/10.1093/protein/gzs025.

73. Fisher MA, McKinley KL, Bradley LH, Viola SR, Hecht MH: **De novo designed proteins from a library of artificial sequences function in *Escherichia coli* and enable cell growth**. *PLoS ONE* 2011, **6**:e15364 http://dx.doi.org/10.1371/journal.pone.0015364.

74. •• Digianantonio KM, Hecht MH: **A protein constructed de novo enables cell growth by altering gene regulation**. *Proc Natl Acad Sci U S A* 2016, **113**:2400-2405 http://dx.doi.org/10.1073/pnas.1600566113.
A non-natural sequence rescues a conditionally lethal gene deletion by modulating the expression of an endogenous promiscuous enzyme.

75. • Digianantonio KM, Korolev M, Hecht MH: **A non-natural protein rescues cells deleted for a key enzyme in central metabolism**. *ACS Synth Biol* 2017, **6**:694-700 http://dx.doi.org/10.1021/acssynbio.6b00336.
A *de novo* protein allows for cell growth by upregulating a metabolic enzyme that can act nonspecifically.

76. •• Donnelly A., Murphy G., Digianantonio K., Hecht M: **A de novo enzyme catalyzes a life-sustaining reaction in *E. coli***. *Nat Chem Biol* 2017 http://dx.doi.org/10.1038/nchembio.2550.
A non-natural sequence enables cell growth by catalyzing enterobactin hydrolysis.

77. Moore J, Arnold F: **Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents**. *Nat Biotechnol* 1996, **14**:458-467 http://dx.doi.org/10.1038/nbt0496-458.

78. Giver L, Gershenson A, Freskgard P, Arnold F: **Directed evolution of a thermostable esterase**. *Proc Natl Acad Sci U S A* 1998, **95**:12809-12813 http://dx.doi.org/10.1073/pnas.95.22.12809.

79. Wintrode P, Miyazaki K, Arnold F: **Cold adaptation of a mesophilic subtilisin-like protease by laboratory evolution**. *J Biol Chem* 2000, **275**:31635-31640 http://dx.doi.org/10.1074/jbc.M004503200.

80. • Smith BA, Mularz AE, Hecht MH: **Divergent evolution of a bifunctional de novo protein**. *Protein Sci* 2015, **24**:246-252 http://dx.doi.org/10.1002/pro.2611.
Directed evolution of a *de novo* protein to strengthen one function resulted in the loss of the unselected function, suggesting that *de novo* proteins can evolve from generalists into specialists.