Quarterly Reviews of Biophysics

cambridge.org/qrb

De novo protein design, a retrospective

Ivan V. Korendovych[1] and William F. DeGrado[2]

[1]Department of Chemistry, Syracuse University, 111 College Place, Syracuse, NY 13244, USA and [2]Department of Pharmaceutical Chemistry and Cardiovascular Research Institute, University of California, San Francisco 555 Mission Bay Blvd. South, San Francisco, CA 94158, USA

Invited Review

Cite this article: Korendovych IV, DeGrado WF (2020). *De novo* protein design, a retrospective. *Quarterly Reviews of Biophysics* **53**, e3, 1–33. https://doi.org/10.1017/S0033583519000131

Received: 17 September 2019
Revised: 13 November 2019
Accepted: 18 November 2019

Key word:
Protein design

Author for correspondence:
William F. DeGrado,
E-mail: William.degrado@ucsf.edu

Abstract

Proteins are molecular machines whose function depends on their ability to achieve complex folds with precisely defined structural and dynamic properties. The rational design of proteins from first-principles, or *de novo*, was once considered to be impossible, but today proteins with a variety of folds and functions have been realized. We review the evolution of the field from its earliest days, placing particular emphasis on how this endeavor has illuminated our understanding of the principles underlying the folding and function of natural proteins, and is informing the design of macromolecules with unprecedented structures and properties. An initial set of milestones in *de novo* protein design focused on the construction of sequences that folded in water and membranes to adopt folded conformations. The first proteins were designed from first-principles using very simple physical models. As computers became more powerful, the use of the rotamer approximation allowed one to discover amino acid sequences that stabilize the desired fold. As the crystallographic database of protein structures expanded in subsequent years, it became possible to construct proteins by assembling short backbone fragments that frequently recur in Nature. The second set of milestones in *de novo* design involves the discovery of complex functions. Proteins have been designed to bind a variety of metals, porphyrins, and other cofactors. The design of proteins that catalyze hydrolysis and oxygen-dependent reactions has progressed significantly. However, *de novo* design of catalysts for energetically demanding reactions, or even proteins that bind with high affinity and specificity to highly functionalized complex polar molecules remains an importnant challenge that is now being achieved. Finally, the protein design contributed significantly to our understanding of membrane protein folding and transport of ions across membranes. The area of membrane protein design, or more generally of biomimetic polymers that function in mixed or non-aqueous environments, is now becoming increasingly possible.

Introduction

The design of small molecules and molecular assemblies with predictable structures has enabled the construction of catalysts, pharmaceuticals, electronics, and smart materials. For example, organic chemists and coordination chemists can design small molecules with well-defined three-dimensional (3D) structures, dynamics, and reactivity. The design of proteins is a much higher order fundamental problem, but one with similarly important implications. It has long been appreciated that the properties of proteins depend on their intricately folded structures. However, we have only recently begun to be capable of designing proteins with predetermined structures. Indeed, 35 years ago it was considered inconceivable that it might ever be possible to design proteins with similar predictability and function.

As in other fields of chemistry, the progress from natural products to fully synthetic proteins has followed a multi-decade path. For example, in the 1950s to 1980s protein drugs, such as insulin and growth hormone were isolated from natural sources. More recently, it has been possible to tap into the immunological repertoire to discover novel antibodies and rationally vary their sequences to create drugs that are addressing multiple unmet medical needs. We are now entering an era in which it has become possible to design proteins with predetermined structures and functions '*de novo*'. This endeavor has already illuminated the principles of protein folding, and proteins are now being designed *de novo* to test and extend our understanding of binding and catalysis.

Here, we discuss the development of *de novo* protein design from its establishment and naming over 30 years ago to early 2019. Before the late 1980s the design of proteins appeared to be impossible. The thermodynamic stability of the native fold of a protein relative to the unfolded form is small and represents the difference between much larger favorable and unfavorable terms, making it very difficult to accurately predict stability. Moreover, the number of possible sequences for even a short protein of 100 residues (20^{100}) is larger than the number of atoms in the universe, precluding the possibility of trying all possible sequences. Indeed, it would not be possible to find a specific sequence by a random search, even if a protein could be mutated every femtosecond for the age of the universe! Similarly, the number of

CAMBRIDGE UNIVERSITY PRESS

possible backbone conformations for a protein of this size represents an astronomically large number (10^{100}), indicating that folding cannot occur by a random search of conformational space (Levinthal, 1969; Bryngelson *et al.*, 1995).

Given the immense complexity of proteins and this prevailing viewpoint, the development of *de novo* protein design was hardly trivial. In its original conception, the *de novo* design of proteins referred to the design of proteins from scratch – rather than by modification of the sequence of naturally occurring proteins (DeGrado *et al.*, 1987; Regan and DeGrado, 1988). It is somewhat surprising that the name has continued to the present, given that W. Feldberg's dictum that a scientist often 'would rather use a colleague's toothbrush than his terminology!' (Katz, 1969). Instead, the meaning of *de novo* design has expanded slightly to include computational methods to redesign natural proteins. *De novo* design also includes sequence-directed approaches, for example, by introducing repeating patterns of apolar and polar residues (DeGrado and Lear, 1985; Kamtekar *et al.*, 1993).

The evolution of *de novo* design occurred in three distinct periods (Table 1). The first era of *manual protein design* using physical models spanned from the late 1970s to the early-1980s. During this period, solid-phase peptide synthesis enabled relatively routine synthesis of peptides up to about 30–50 residues in length. However, gene synthesis was not yet routine, limiting the size of proteins that could reliably be produced. The second wave, spanning from the mid-1980s to the early 2000s focused on *computational design guided by fundamental physicochemical principles*. Proteins were designed using mathematical equations to define the backbone conformations (DeGrado *et al.*, 1987; Regan and DeGrado, 1988; Harbury *et al.*, 1995) and sidechain repacking algorithms to design the sequence (Ponder and Richards, 1987; Desjarlais and Handel, 1995; Dahiyat and Mayo, 1996). This period also marked the first example of cooperatively folded proteins (DeGrado *et al.*, 1987; Regan and DeGrado, 1988), the first computationally repacked natural protein domains (Dahiyat and Mayo, 1997; Lazar *et al.*, 1997), and the first computationally designed completely *de novo* protein whose structure was fully verified (Walsh *et al.*, 1999). The third wave began in the early 2000s as the expanding database of crystallographic structures enabled *fragment-based and bioinformatically informed computational protein design*. The Protein Data Bank (PDB) was deconstructed into a list of parts consisting of protein fragments, each with defined sequence preferences and interaction patterns that could be reassembled to create novel proteins (Kuhlman *et al.*, 2003; Huang *et al.*, 2016*a*).

Table 1 highlights a number of key advances from the first 20 years, up to the development of fragment-based design of a protein designated TOP7, which was accomplished in 2003 (Kuhlman *et al.*, 2003). Beyond this point, the field expanded rapidly, and the accomplishments are too many and varied to easily tabulate. Today, protein designers combine the essential tools from each of these periods. *De novo* design has already passed a number of milestones, the first of which was the construction of sequences that folded in water and membranes to adopt precisely predetermined folded conformations. Complex functions have also been achieved, ranging from binding and catalysis to transmembrane (TM) ion and electron transport. Here, we focus on the original question posed by the field of *de novo* design, is our knowledge of the principles of folding and function sufficient to design proteins from scratch. Therefore, we focus almost exclusively on *de novo* proteins whose structures and sequences have been designed using a mathematical parameterization or fragment assembly, rather than using the sequences or 3D structures of natural proteins as the starting point. To maintain this focus we do not discuss combinatorial sequence-based approaches such as binary patterning. We instead refer the reader to reviews of this outstanding work (Hecht *et al.*, 2004, 2018). Also, wherever possible, we restrict our discussion to proteins whose structures and/or dynamics have been very extensively characterized by high-resolution methods.

Manual protein design

As early as 1979, Bernd Gutte used manual model building and physical models to design a 35-residue RNA-binding peptide (Gutte *et al.*, 1979), followed by a 25-residue peptide intended to bind dichlorodiphenyltrichloroethane (DDT) in 1983 (Moser *et al.*, 1983) (Fig. 1). While some binding was observed, solubility problems precluded determination of their structures. In the mid-1980s, Jane and David Richardson began their collaborative work with Bruce Erickson aimed at the design of 'betabellins' (and the related 'betadoublets'), meant to mimic the structure of β-sandwich proteins (Richardson and Richardson, 1989). In this case, computer graphics and secondary structure propensities gleaned from analysis of natural proteins were used to facilitate the design process. Again, poor solubility and aggregation proved to be problematic. Ultimately, Erickson demonstrated that at least one member of this class of designed proteins formed amyloid-like fibrils (Lim *et al.*, 1998). In retrospect, it is likely that the formation of amyloid-like structures explained the ability of Gutte's DTT-binding peptides to bind hydrophobic substances (West *et al.*, 1999). A variety of amyloids are well known to bind a variety of flat-aromatic molecules including amyloid dyes (West *et al.*, 1999). Attempts to increase solubility and decrease aggregation of the betabellin and betadoublet families of proteins led to derivatives with fluctuating structures that defied high-resolution structure determination (Quinn *et al.*, 1994). The design of uniquely folded β-proteins continued to be challenging, and accurate design of such tertiary structures was achieved only in the last two years (Dou *et al.*, 2018).

Thus, by the mid-1980s, although there were sporadic attempts to design proteins with predetermined structures and functions, this goal had not been achieved. However, this was about to change due to a number of concurrent technical advances.

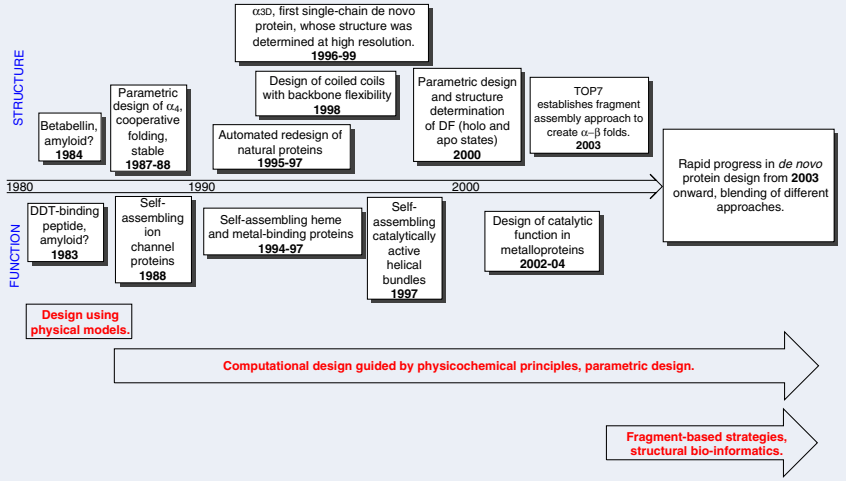
Computational design guided by fundamental physicochemical principles

Helical bundles, the first structurally defined proteins designed from scratch

In the 1970s and 1980s, a number of key advances made *de novo* protein design feasible for the first time. Methods of solid phase peptide synthesis had reached an advanced stage for the synthesis of sequences up to about 50 residues, and the synthesis of synthetic genes had become increasingly possible, allowing one to design larger novel proteins. Computer graphics coupled with methods of molecular mechanics and dynamics allowed one to work with highly complex structures, freeing the designer from working with cumbersome physical models. Crystallographic and nuclear magnetic resonance (NMR) methods were also rapidly improving. Finally, site-directed mutagenesis of natural proteins provided a better understanding of the energetics and kinetics of protein folding – and the contributions of individual

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

Table 1. The formative first 20 years of *de novo* protein design 1983–2003



Year	Key contributors and references	Accomplishment
The first wave of protein design, beginnings		
1983	Bernd Gutte *et al.* (Moser *et al.*, 1983)	Design of a 25-residue DDT-binding peptide. Showed β-sheet secondary structure based on CD, likely amyloid-like.
1984	Jane and David Richardson with Bruce Erickson (Unson *et al.*, 1984)	Design of betabellin, a covalent homodimer of a 32-residue peptide. Although water solubility and amyloid formation were initially problematic, a more soluble peptide with the appropriate secondary structure was subsequently achieved. However, the structures remained too dynamic to be amenable to structure determination (Quinn *et al.*, 1994).
The second wave: Computational design guided by fundamental physicochemical principles		
1987–88	DeGrado *et al.* (DeGrado *et al.*, 1987; Ho and DeGrado, 1987; Regan and DeGrado, 1988)	Water-soluble protein design: a parametric approach was used to achieve the first cooperatively folded, stable, water-soluble protein, α4. Its tertiary structure appeared approximately correct based on introduction of disulfides and Zn(II)-binding sites.
1988	Lear, Wasserman, and DeGrado (Lear *et al.*, 1988; DeGrado *et al.*, 1989)	Design of TM α-helices that assemble into ion channels in membranes. This work helped define key features required for helical assembly in membranes and ion conduction.
1994–1997	Pecoraro, DeGrado, and Dutton (Choma *et al.*, 1994; Robertson *et al.*, 1994; Dieckmann *et al.*, 1997)	Self-assembling peptides that bind metal ions and hemes.
1995–97	Desjarlais & Handel, Dahiyat & Mayo (Desjarlais and Handel, 1995; Dahiyat and Mayo, 1996; Dahiyat and Mayo, 1997; Lazar *et al.*, 1997)	Automated redesign of protein sequence achieved by repacking the backbone structures of small domains of natural proteins.
1996–1999	DeGrado *et al.* (Betz *et al.*, 1996; Walsh *et al.*, 1999)	Successful design of α3D, the first computationally designed *de novo* globular protein whose complete structure was solved at high resolution.
1997	Baltzer (Broo *et al.*, 1997; Broo *et al.*, 1998)	Design of catalytically active helix–loop–helix motif that assemble into four-helix bundles.
1998	Harbury, Kim and Alber (Harbury *et al.*, 1998)	Design of coiled coil using flexible-backbone methods and parametric equations.
2000–04	Lombardi and DeGrado *et al.* (Lombardi *et al.*, 2000*b*; Summa *et al.*, 2002; Kaplan and DeGrado, 2004)	The use of parametric equations to design the first *de novo* metalloprotein (whose structures were determined at high resolution), including catalytically active metalloproteins with extensive buried hydrogen-bond networks that mediate assembly and interaction specificity.
The third wave: fragment-based and bioinformatically informed computational protein design		
2003	Kuhlman and Baker (Kuhlman *et al.*, 2003)	Use of fragment build-up procedures to design TOP7, the first globular protein to be designed with a fold not observed in nature.

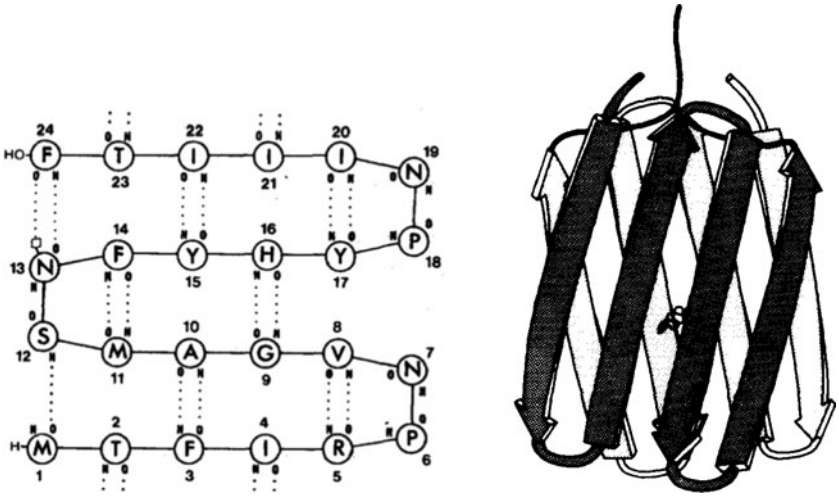


Fig. 1. (Left) Proposed secondary structure of a DDT-binding peptide (reproduced with permission from Moser *et al*. (1983)). (Right) Molecular model of a short segment of the amyloid fibril formed by betabellin (reproduced with permission from Richardson and Richardson (1989)).

sidechains to the process. It became generally accepted that the packing of hydrophobic residues in the solvent-accessible interiors of proteins contributed significantly to the driving force for water-soluble protein folding, and that polar interactions, although less favorable, often helped define the detailed geometries of protein structures (Fersht and Serrano, 1993). Moreover, the preferences of amino acids for adopting specific secondary structures and rotamers enabled computational methods to select a sequence to stabilize a given fold (Box 1).

Thus, by the 1980s the stage was set for *de novo* protein design. Nevertheless, there was considerable skepticism that *de novo* design would be possible given the astronomical number of potential sequences and conformations for even a modestly long protein sequence. How then, might proteins have evolved within the first billion years after the formation of our planet? One attractive hypothesis was that modern-day proteins evolved from self-association of short peptides capable of forming secondary structural or other functional units. Dayhoff suggested that structures could be assembled through intermolecular association of multiple chains or by intramolecular association (folding) of proteins formed by duplicating of genes expressing for the primordial units (Eck and Dayhoff, 1966). DeGrado and Lear (1985) hypothesized that some of the first precursors to natural proteins were amphiphilic peptides, in which hydrophobic and polar residues segregate on opposite sides of an α-helix or β-sheet; assembly of the hydrophobic faces would drive folding in an aqueous environment. To test this hypothesis, they designed peptides composed of only Leu and Lys as hydrophobic and polar residues. When the polar and apolar residues were alternated in the sequence to match the geometric repeat of the β-sheet, the resulting peptide (LKLKLKL) assembled into a β-conformation in aqueous solution. However, when the polar and apolar residues were allowed to match that of an α-helix in $(LKKLLKL)_2$, the peptides self-associated into tetrameric bundles of α-helices, which the authors speculated might have 222 symmetric structures similar to the recently recognized family of natural antiparallel four-helix bundle folding motif (Fig. 2*a* and *d*) (Argos *et al*., 1977; Weber and Salemme, 1980; Presnell and Cohen, 1989; Beesley and Woolfson, 2019).

This investigation set off a series of studies that culminated in the design of large families of helical bundle proteins. Success in designing a protein that folded into a desired structure did not come immediately, but instead in stages, as we came to understand the requirements for secondary structure formation, folding into a globular thermodynamically stable ensemble of closely related proteins, and ultimately into a single well-folded protein structure. Early attempts to crystallize $(LKKLLKL)_2$ in the lab of David Eisenberg were unsuccessful. Therefore, DeGrado and Eisenberg collaborated on the redesign of the sequence of $(LKKLLKL)_2$ to better stabilize the desired antiparallel tetrameric structure.

The initial design idealized the approximate D_2 symmetry of natural four-helix bundles (Eisenberg *et al*., 1986) (i.e. with two-fold rotational symmetry axes running down the bundle as well as between neighboring antiparallel helices, labeled Z, Y, and X in Fig. 2*a*) of natural four-helix bundle proteins. Internal symmetry reduced the size of the sequence space that needed to be considered, and it allowed the basic unit to be used repeatedly to build the entire four-helix bundle. Similar parametric models with minimal numbers of adjustable parameters (Salemme, 1983; Lasters *et al*., 1988; Betz and DeGrado, 1996; Lombardi *et al*., 2000*b*; North *et al*., 2001; Offer *et al*., 2002; Emberly *et al*., 2004; Grigoryan and DeGrado, 2011) have since been used to build-up more complex tertiary structures including the rubredoxin fold (Nanda *et al*., 2005), Triosephosphate isomerase (TIM) barrels (Huang *et al*., 2016*b*), β-barrels (Dou *et al*., 2018), β-propellers (Voet *et al*., 2014), coiled coils (Harbury *et al*., 1998; Huang *et al*., 2014; Thomson *et al*., 2014), and repeat proteins (Brunette *et al*., 2015), as discussed in subsequent sections. In parametric protein design, one begins with highly symmetrical backbones, to create a 'draft' of the desired structure and then lifts symmetry as needed to accommodate the asymmetric placement of loops and active sites (Lombardi *et al*., 2000*b*; Huang *et al*., 2016*a*; Polizzi *et al*., 2017; Dou *et al*., 2018).

The designed self-associating tetrameric peptide, α1A was built manually using a set of physical 'Kendrew' models (Eisenberg *et al*., 1986). Physicochemical principles guided all aspects of the design. Leu sidechains were chosen for the hydrophobic interior, where they were able to interdigitate in low-energy rotamers. Helix-promoting Glu and Lys residues were chosen for the exterior-facing residues, and they were arranged to form favorable electrostatic interactions. Although α1A failed to crystallize (Ho and DeGrado, 1987), a short, 12-residue fragment of α1A (designated α1) isolated as a byproduct of the synthesis was crystallized. Too short to form the desired full-length bundle, this peptide assembled into

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

Box 1. Setting the stage for *de novo* protein design, sidechain packing algorithms, and automated sequence selection

Early studies showed that the sidechains in protein cores adopted low-energy conformations called rotamers (Janin *et al*., 1978), which were tightly packed with an efficiency similar to small molecule crystal lattices (Richards, 1977). The distribution of each rotamer was subsequently shown to depend on secondary structure (McGregor *et al*., 1987; Dunbrack and Karplus, 1993; Dunbrack and Cohen, 1997). These findings led to a model in which sidechains were packed in protein cores as in a 3D jigsaw puzzle. These two requirements – that side chains form stable rotamers and that they be efficiently packed in protein interiors – provided two powerful restraints that define the interior-facing residues of uniquely folded globular proteins. The first cooperatively folded globular *de novo* proteins were designed following these imperatives by using minimal set of apolar and polar sidechains (DeGrado and Lear, 1985; Eisenberg *et al*., 1986; DeGrado *et al*., 1987; Ho and DeGrado, 1987).

As computational power increased it became possible to consider the repacking protein of cores with the full set of natural amino acids. Here, one begins with a given backbone structure and explores large numbers of side chains that can fit together to stabilize the fold (Ponder and Richards, 1987). Ideally, each possible combination of sidechain and rotamer identities would be evaluated at each position, but the number of combinations rapidly becomes unmanageable without the use of computational algorithms, including genetic (Jones, 1994; Willett, 1995), Monte-Carlo (Metropolis *et al*., 1953), and dead-end-elimination (Desmet *et al*., 1992; Lasters *et al*., 1995; Gordon *et al*., 2003) algorithms. In 1987, Ponder and Richards introduced sidechain repacking algorithms to probe the combinatorics of packing in natural proteins (Ponder and Richards, 1987). In 1995, Desjarlais and Handel (Desjarlais and Handel, 1995; Johnson *et al*., 1999) used repacking algorithms together with a genetic algorithm to redesign the core of small natural protein domains. In a series of landmark papers (Dahiyat and Mayo, 1996; Dahiyat and Mayo, 1997), Mayo and coworkers expanded repacking algorithms to include selection of exterior sidechains, as well as the use of dead-end-elimination to facilitate the search. In 1997, Dahiyat and Mayo achieved the completely automated redesign of the sequence of a natural 28-residue Zn(II) finger motif peptide, starting with only the backbone structure of the second zinc finger module of the DNA binding protein Zif268. (Dahiyat and Mayo, 1997). In the same year, Handel, Desjarlais, DeGrado, and coworkers introduced sidechain repacking algorithms to design a protein whose backbone was not taken from a natural protein. The structure of the resulting 73-residue protein, α3D was in excellent agreement with the design (Betz *et al*., 1996; Bryson *et al*., 1998; Walsh *et al*., 1999). Today, sidechain repacking algorithms represent an important part of all fully atomistic computational approaches to protein design.

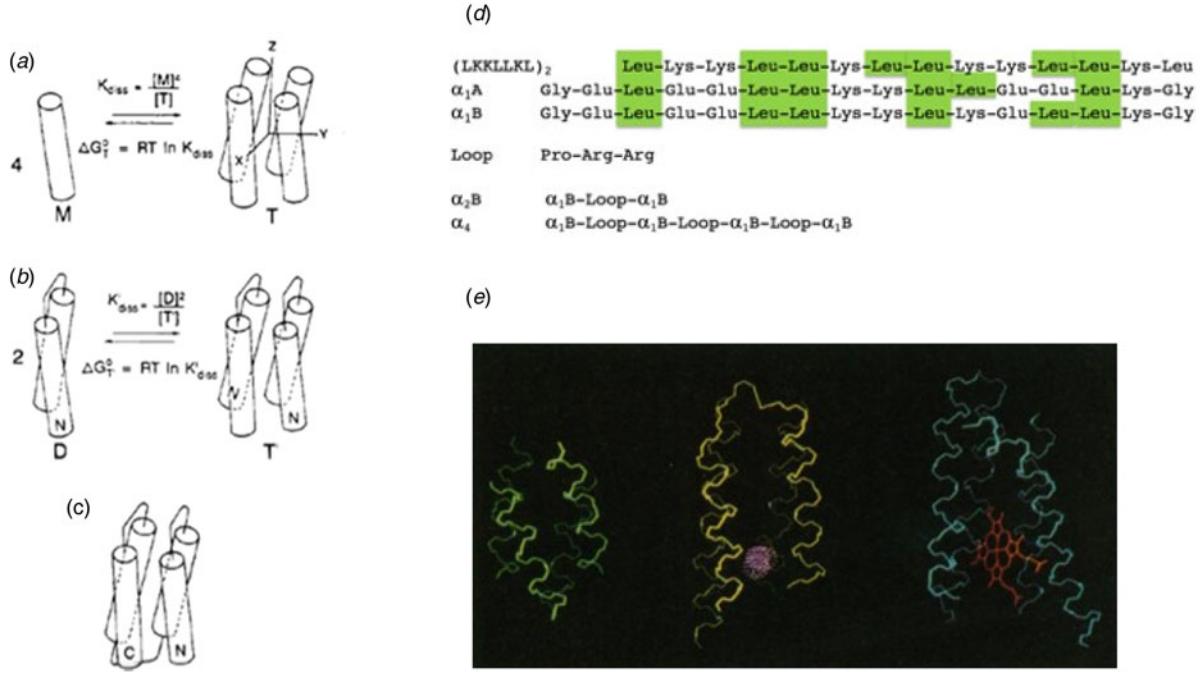


Fig. 2. Design of a four-helix bundle. (*a*) A peptide was designed, which self-associated to form an antiparallel helical bundle in solution. A loop sequence was next inserted (*b*) between two helices to create a dimeric four-helix bundle, and then three loops were inserted between four helices to create the full-length helical bundle. At each stage, the free energy of assembly or folding was determined, and used to evaluate possible sequences. In this way, the complex problem of protein design was cut into smaller separable pieces. For simplicity, the monomeric species in panels (*a*) and (*b*) are shown as helices, but they were actually only partially helical, as shown by CD. Panel (*d*) shows the sequences of the peptides and proteins discussed in the text. Panel (*e*) shows an early energy-minimized model of α4 (left) as compared to larger natural four-helix bundle proteins (myohemerythrin, middle) and cytochrome c′ (right). Panels (*a*–*c*) are reproduced with permission from Ho and DeGrado (1987). Copyright (2007) American Chemical Society, while panel (*e*) is reproduced with permission from DeGrado *et al*. (1989).

multiple association states in solution and the solid state (Patterson *et al*., 1999; Prive *et al*., 1999).

Ho and DeGrado (1987) next used computer graphics and energy minimization to redesign the sequence, minimizing the exposure of apolar residues on the surface. In contrast to α1, the resulting full-length α1B peptide cooperatively assembled into a highly stable tetrameric four helix bundle (-22 kcal mol^{-1}, 1 M standard state). The α1B tetramer was compact and globular,

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

and detailed NMR investigations also showed that the helices began and ended precisely as in the design (Osterhout *et al.*, 1992). The first attempt to build loops between the helices revealed an important and previously unarticulated aspect of protein folding – the sequence of a protein must not only stabilize the desired fold. Instead it must destabilize all closely related folds while stabilizing the native structure (DeGrado *et al.*, 1987; Ho and DeGrado, 1987).

The final α4 protein was 74 residues in length, and expressed well in bacteria. It represented the first example of a *de novo* designed protein with a cooperatively folded, globular conformation in aqueous solution (DeGrado *et al.*, 1987; Regan and DeGrado, 1988). Furthermore, it was highly stable, with a cooperative equilibrium unfolding transition near 6 M guanidine hydrochloride. Clearly, the first milestones in *de novo* protein design had been passed. Furthermore, structure-stabilizing disulfides (Regan *et al.*, 1994) and metal-binding sites (Handel and DeGrado, 1990; Regan and Clarke, 1990; Handel *et al.*, 1993) were successfully introduced into the tertiary structure, as confirmed by NMR (Handel and DeGrado, 1990; Handel *et al.*, 1993). Thus, the Zn^{2+}-binding derivatives of α4 indeed achieved the correct overall fold that positioned residues distant in sequence into close proximity to create the functional binding site. A second milestone was crossed.

Over the past few decades, studies of natural proteins have shown that they can natively achieve a wide-ranging spectrum of order, ranging from intrinsically disordered (random coil), to compact but flexible, to ones with well-packed cores. However, in the 1980s there was less understanding of this spectrum of native states, so there was considerable interest in determining the degree of structural uniqueness that could be achieved with a minimal protein such as α4. Solution NMR and fluorescence studies showed that the buried hydrophobic residues of α4 were conformationally more mobile than those of most crystallographically characterized proteins. Over the next decade, various groups attempted to address this issue, as reviewed previously (Bryson *et al.*, 1995; DeGrado *et al.*, 1999), and only a few early contributions will be mentioned here. Expecting that a more diverse sequence might lead to improved properties, Jane and David Richardson designed a protein, called FELIX, which incorporated all the natural amino acids (Hecht *et al.*, 1990). However, FELIX had very marginal stability (around -1 kcal mol^{-1} *versus* -20 kcal mol^{-1} for α4), and subsequent studies by this group showed that it did not unfold in a cooperative transition – instead they concluded that FELIX adopted a 'non-stable and non-unique tertiary structure' (Gernert *et al.*, 1993). Stroud *et al.* constructed a monomeric four-helix bundle by stitching loops between four identical helical peptides (Schafmeister *et al.*, 1997) that had originally been designed to solubilize membrane proteins, but instead were found to self-associate into a tetrameric four-helix bundle (Fig. 3*a*). Although a crystal structure was determined, the loops were disordered, so it was not possible to determine the topology of the bundle. Finally, by introducing polar interactions and introducing geometric complementarity into the originally designed α2B scaffold, it was possible to design and structurally characterize uniquely folded dimeric four-helix bundles (Hill and DeGrado, 1998, 2000; Hill *et al.*, 1999, 2000).

A breakthrough in *de novo* design of uniquely folded proteins occurred with the ability to computationally 'repack' the hydrophobic core of designed backbones (Ponder and Richards, 1987). As mentioned above, Handel (Desjarlais and Handel, 1995) and Mayo (Dahiyat and Mayo, 1997) demonstrated the use of these algorithms for repacking the core of small natural protein domains. DeGrado, Handel, and coworkers introduced the use of these algorithms to design of a *de novo* protein (Betz *et al.*, 1996; Bryson *et al.*, 1998), rather than starting with the 3D structure of a natural protein. They designed a three-helix bundle, α3D, through sidechain repacking and energy minimization. The interior sidechains consisted of a diverse set of apolar residues that packed in a geometrically complementary manner. Interhelical electrostatic interactions at solvent-exposed positions were also used to specify a single topology. The NMR structure (Fig. 3*b*) (Walsh *et al.*, 1999) was in close agreement with the design providing the first example of the *de novo* design of a globular protein with an accurately predetermined structure. Another important milestone in *de novo* design had been passed. Given its relatively simple but cooperatively folded globular structure, α3D quickly became a very widely studied protein for computational and experimental studies of protein folding (Zhu *et al.*, 2003; Park *et al.*, 2006; Liu *et al.*, 2009; Adhikari *et al.*, 2012; Shao, 2014; Chung *et al.*, 2015; Zeng *et al.*, 2016; Maruyama and Mitsutake, 2017; Walder *et al.*, 2017; Xiong *et al.*, 2017; Jumper *et al.*, 2018; Koebke *et al.*, 2018; Yoo *et al.*, 2018; Gadzala *et al.*, 2019). Its folding kinetics are among the most extensively characterized of small cooperatively folded proteins (Chung *et al.*, 2015). The protein α3D has also become as a template for the design of metalloproteins (Fig. 3*c*) (Chakraborty *et al.*, 2011; Mocny and Pecoraro, 2015; Tebo and Pecoraro, 2015; Plegaria and Pecoraro, 2016). Many examples of functional helical bundles based on α3D and designed four-helix scaffolds were soon to follow, as discussed below.

The design of uniquely folded proteins also coincided in time with the understanding that proteins fold in a funnel-like manner, accruing increasing native tertiary structure as folding progresses. This smooth process is known as minimal frustration (for a review see, Wolynes (2015)). The final ensemble of states – whether it be a uniquely and tightly packed 3D structure or a more loosely folded 'molten globule' – depends on whether the sequence can assume one single backbone structure and sidechain packing arrangement or a more energetically diverse set of structures and packings. One of the surprises of protein design was that the folding landscape can so easily occur with minimal frustration, and that consideration of the native state frequently leads to a foldable sequence that does not get 'stuck' in numerous off-pathway solutions. The smoothness of the folding funnel for natural proteins has often been discussed in terms of evolution. We believe it is also an intrinsic propensity of the properties and geometry of the protein backbone and the reliance on the hydrophobic interaction to drive folding in nature. The need to tightly pack the amide backbone leads to highly compact secondary structures in which the polar amides form intramolecular hydrogen bonds to compensate for stabilizing interactions with water that occur in the unfolded state. Ultimately, the burial and favorable packing of apolar sidechains in the protein interior drives folding, while the requirement to maintain water-solubility dictates the predominant placement of polar residue on the exterior. Together, these restraints lead not only to a stable folded structure, but also to minimal frustration along the folding pathway. It is also interesting to note that the misfolding of the same protein sequences into amyloids (Chiti and Dobson, 2009) generally occurs through aggregation, presumably on a much rougher landscape. The ability to fold rapidly along a smooth funnel (and hence kinetically escape amyloid formation in a non-equilibrium

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

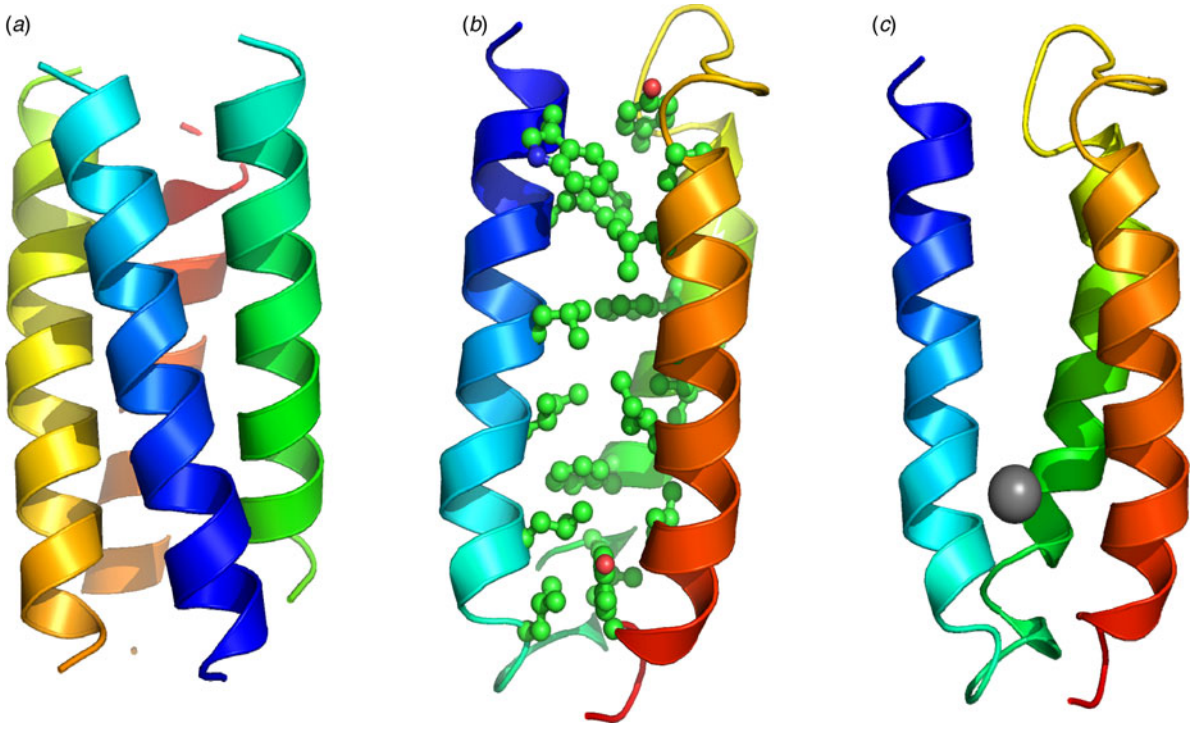


Fig. 3. (*a*) Crystal structure of a peptide that was designed to solubilize membrane proteins, but was serendipitously found to crystallize as a four helix coiled-coil bundle DHP1 (PDB: 4HB1). (*b*) NMR structure of α3D (PDB: 2A3D) is stabilized by a set of apolar sidechains that pack in a geometrically complementary manner, shown in ball-and-stick format. (*c*) The model of 3-His α3D based on EXAFS data and NMR structure of α3D (PDB: 2A3D).

living system) must have been one of the earliest features in the molecular evolution of proteins.

Coiled coils

Coiled coils represent a special class of helical bundles, which have been particularly useful stepping stones in the development of *de novo* protein design. The α-helical coiled coil (Fig. 4) represents a structure of intermediate complexity, bridging the gap between simple monomeric helices and native proteins. The classical left-handed coiled-coil has a seven-residue geometric repeat labeled, '*abcdefg*'; '*a*' and '*d*' side-chains project toward the bundle core and are mostly hydrophobic whereas '*e*' and '*g*' residues face the inter-subunit interface and are generally more polar (Crick, 1953). Hodges and co-workers used a sequence-based approach to design repeating heptapeptides as models for two-stranded coiled coils. In the prototype, $(Leu_aGlu_b,Ala_cLeu_dGlu_eGly_fLys_g)_n$, apolar Leu residues at positions '*a*' and '*d*' of the heptad hydrophobically stabilize the structure (Lau *et al.*, 1984). This heptad repeat formed the basis for the design of a 29-residue peptide (O'Neil and DeGrado, 1990) that was used to determine the helical propensities of various amino acids. Subsequent determination of the crystal structure of this peptide showed that it formed a trimeric antiparallel structure, rather than the expected parallel dimer. Shortly thereafter, studies on derivatives of the two-stranded coiled-coil domain of a yeast transcription factor, GCN4, further illustrated the role of polar and packing interactions in determining the stoichiometry and topology of coiled coils (Harbury *et al.*, 1993, 1994). Alber, Harbury, Kim, and coworkers showed that van der Waals (vdW) packing between buried residues at the '*a*' and '*d*' positions play critical roles in determining the stoichiometry and structure of coiled coils. Amino acid substitutions as subtle as Leu-to-Ile substitutions switch the assembly from favoring trimers to tetramers, and this switch could be understood and predicted based on simple packing arguments. Moreover, Alber, Harbury, and Kim introduced the use of flexible-backbone methods and parametric equations to design both right-handed and left-handed coiled coils (Harbury *et al.*, 1998), representing another important milestone in *de novo* protein design.

More recently, Woolfson and coworkers extended these studies to the design 4- to 8-stranded bundles by manipulating the physicochemical and steric properties of the residues at the '*e*' and '*g*' positions (Fig. 4) (Thomson *et al.*, 2014). Importantly, coiled coils with some of these association states had never been characterized before – yet another milestone in *de novo* protein design. Moreover, Baker and coworkers extended the use of parametric equations to design regular bundles, with a variety of geometric repeats and stoichiometries (Huang *et al.*, 2014). They also automated the process of searching for backbones that allow the formation of hydrogen-bond networks into homo- and heterodimeric coiled coils (Boyken *et al.*, 2016; Chen *et al.*, 2019). Today, the design of regular coiled coils of various sizes and shapes would appear to be a solved problem.

Functional de novo *designed helical bundles*

As the principles for designing structurally unique helical bundles became better understood, it also became possible to design functions. The interior of helical bundles can be elaborated to bind a variety of metal ions and small substrates. Much of this work predated the development of integrated packages for protein structure prediction and design such as Rosetta, and instead relied on physical principles and molecular mechanics force fields to

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

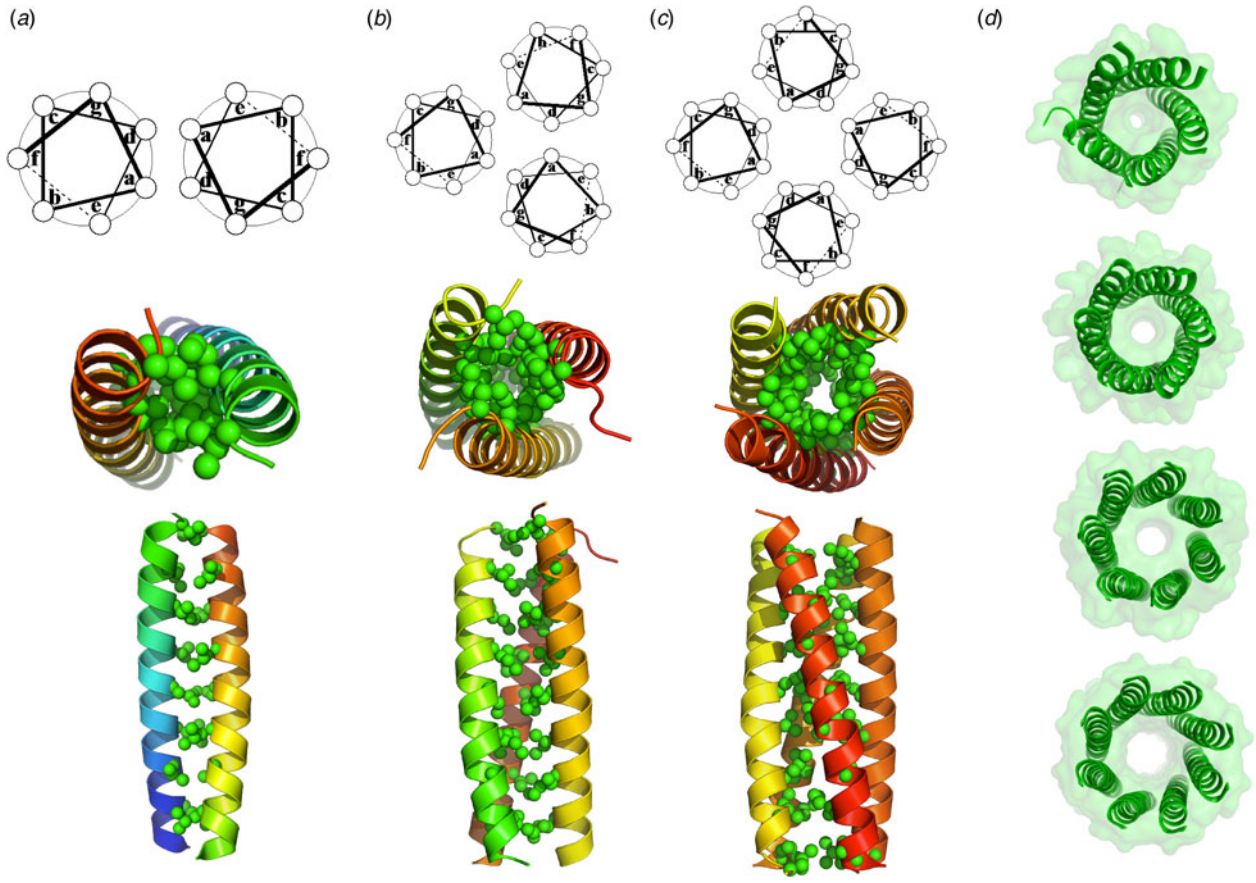


Fig. 4. (*a*) A crystal structure of a dimeric natural coiled-coil GCN4 interaction (PDB: 2ZTA) and the corresponding helical wheel. (*b*) A side on and end on views of the hydrophobic interior of a trimeric coiled-coil GCN4 derivative (PDB: 1GCM) along with the corresponding helical wheel. (*c*) A side on and end on views of the hydrophobic interior of a tetrameric GCN4 derivative (PDB: 1GCL) along with the corresponding helical wheel. (*d*) End on views of *de novo* designed penta-, hexa-, hepta-, and octameric bundles (PDB: 4PND, 4H8O, 5EZ8, 6G67).

guide the designs. More recently, Rosetta has brought most of the essential steps into a single framework, simplifying the overall process and allowing inclusion of structural bioinformatics data into the design process (Leaver-Fay *et al.*, 2011).

Overall strategy for building metal ion and cofactor-binding sites

Metal ion sites in proteins serve both structural and functional roles. Structural sites, such as in zinc fingers, tend to have common, coordinately saturated geometries that stabilize the folded conformation of the protein. In contrast, functional sites often have coordinately unsaturated in geometries that are enforced by the fold of the protein. Metalloproteins catalyze a remarkable array of reactions and a given metal ion such as manganese or iron can be used in different enzymes to catalyze a number of oxidative, reductive, and hydrolytic transformations (Yu *et al.*, 2014). Thus, the activity of a given metalloprotein represents a partnership between the metal ion cofactor and the protein matrix: the metal ion brings non-discriminate chemical reactivity, while the protein stabilizes the metal ion in aqueous solution, fine tunes its reactivity, and binds substrates for catalysis. The protein also often positions hydrogen bond donors, acceptors and tunes the electrostatic environment for catalysis. *De novo* design allows us to probe and expand our understanding of these processes.

While it is possible to graft metal ion sites into existing proteins, in our approach to *de novo* design of metalloproteins the geometrically stringent requirements for metal ion and substrate binding instead dictate the backbone of the protein (Lombardi *et al.*, 2000*b*). The ligation geometry and the requirement that the ligating sidechains adopt energetically accessible conformations together provide powerful restraints that help define the overall fold and backbone structure. Second-shell hydrogen bonds to the primary ligands provide an additional restraint, which further restricts the possible backbone geometries. The function dictates the nature of the ligands (most commonly, Met, Cys, Asp/Glu, and His) employed in a given design. The nature of the ligands and their geometry help control the affinity and redox properties of the bound metal ion as well as its Lewis acidity. The availability of ligation sites for interaction with exogenous ligands, including water, O_2, and organic substrates provides another important restraint. Finally, flexibility must be considered to stabilize multiple states as substrates come on and off, and, in some cases, the metal ions change oxidation state.

When the site is symmetrical this can facilitate parametric design of the protein backbone as illustrated in Fig. 5. The design is completed by introduction of loops, and sequence selection completed as in the above section. As described below, while the initial designs are often symmetrical, it is frequently necessary to lift the symmetry in subsequent designs as required for function.

Di- and tetranuclear metal complexes

Dimetal (e.g. di-Co, di-Fe, and di-Mn) proteins catalyze a variety of hydrolytic and redox processes (Marsh and Waugh, 2013;

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

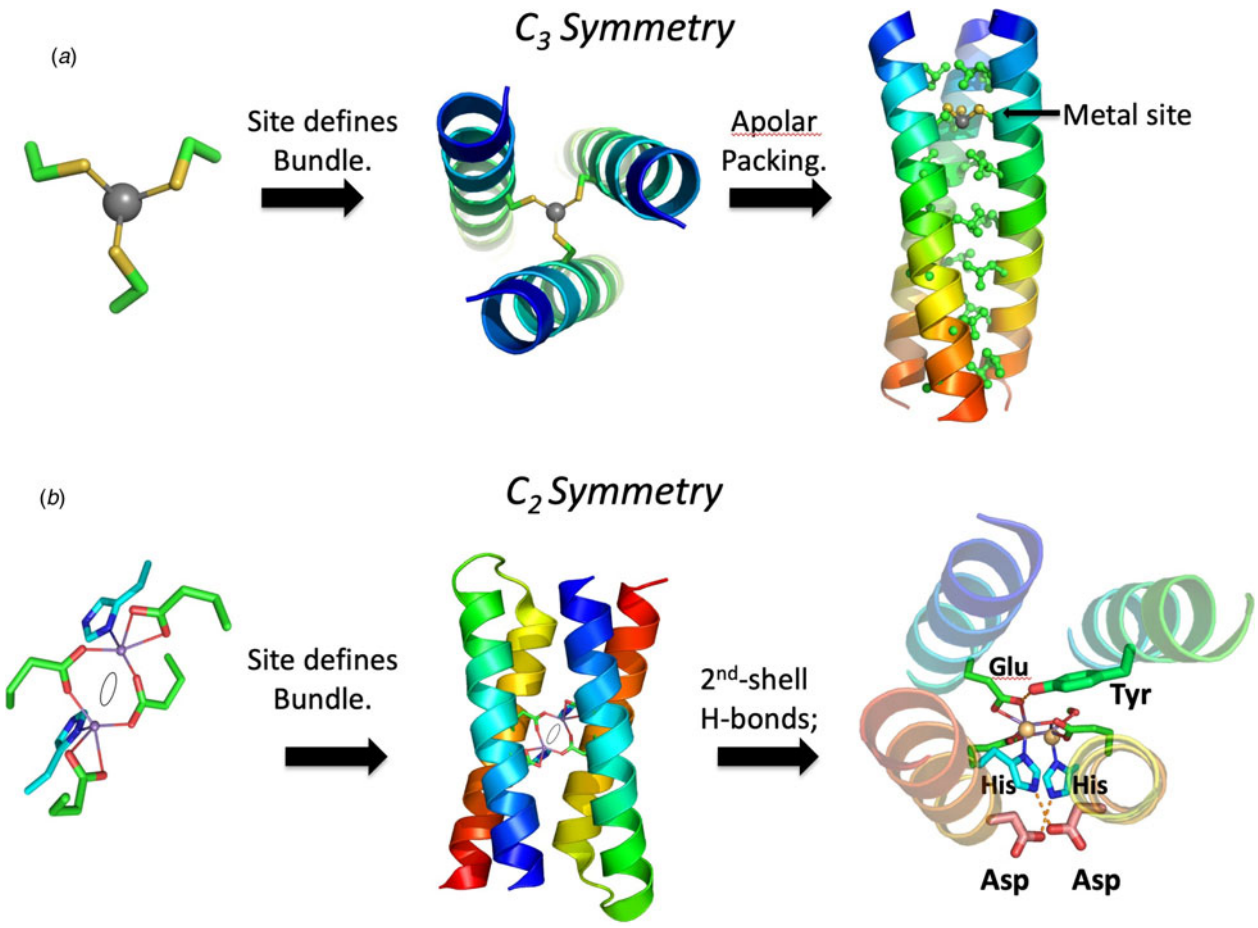


Fig. 5. The desired geometry of the metal ion-binding site dictates the overall 3D structures during *de novo* protein design. In panel (*a*), a trigonal 3-Cys site dictates the backbone of a three-helix bundle in the TRI series of peptides (Dieckmann *et al*., 1997, 1998; Mocny and Pecoraro, 2015) (PDB: 2JGO). The structure is stabilized in the desired conformation by favorable vdW packing and the hydrophobic interactions between buried apolar residues (far right). In panel (*b*), a more complex C_2 symmetrical site is formed from 4-Glu and two-His residues, which bind to two transition metal ions in a four-helix bundle in the DF series of proteins (Lombardi *et al*., 2019). The two-fold axis is denoted by an oval. A large number of second-shell hydrogen bonds were positioned to stabilize the ligands in the desired conformation, and the remaining interior residues chosen (not shown) were apolar sidechains that pack efficiently in the interior of the bundle.

Wang *et al*., 2015; Jasniewski and Que, 2018; Crichton, 2019). Their metal-binding sites are rich in Glu/Asp and His ligands, and the metal ions are generally bridged by water (also OH^- or O^{2-}) and/or carboxylate-containing sidechains. We were particularly drawn to the O^{2-} utilizing proteins, which include hydroxylases, fatty acid desaturases, radical-generating ribonucleotide reductases, catalases, ferritins, and aldehyde decarbonylases. Although the overall structures of these proteins are highly diverse, in each case the di-Mn or di-Fe sites of all these proteins are housed within an antiparallel four-helix bundle that is generally embedded into a much larger structure (Summa *et al*., 1999; Lombardi *et al*., 2000*b*).

In 2000, Lombardi and DeGrado designed a minimal diiron protein (DF) (Lombardi *et al*., 2000*b*), not by modification of the sequence of a natural diiron protein, but rather by starting from first-principles and using a set of equations to generate the fold of the structure. The backbone was a D_2-symmetric four-helix bundle – each helix donating a single Glu ligand. An additional His residue was placed on just two of the helices, leaving two free sites to interact with substrates such as O_2. The final model was a two-fold symmetric dimer of helical hairpins, whose backbone structure was dictated by: (1) coordination requirements of the Glu_4His_2-diiron site; (2) suitable helical packing angles and distances; and (3) Asp and Tyr second-shell H-bonds to the coordinating His and Glu (Fig. 5). The core was packed using the algorithm of Desjarlais and Handel (1995).

Remarkably, the first designed sequence folded into a very stable dimetal-binding protein; for the first time, a *de novo* metalloprotein showed a crystal structure in excellent agreement with the intended design (Lombardi *et al*., 2000*b*). Both the backbone and the entire network of first- and second-shell ligands were realized precisely as in the intended design (Figs 6*a* and *b*). Moreover, the solution NMR structure of metal-free *apo*-DF1 was nearly identical to the holo-protein, indicating that the six coordinating and the four second-shell ligands were largely preorganized with Å-level accuracy even in the absence of the metal cofactor (Maglio *et al*., 2003). Thus, DF imposed its structure onto the metal cofactor rather than *vice versa*, demonstrating that a preorganized binding site in the apolar core could be stabilized by a sufficient set of H-bonds and salt bridges.

In subsequent work, DF1 was engineered to realize a number of binding and catalytic functions. Each step illustrated a tradeoff between protein stability and function (Shoichet *et al*., 1995). The

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

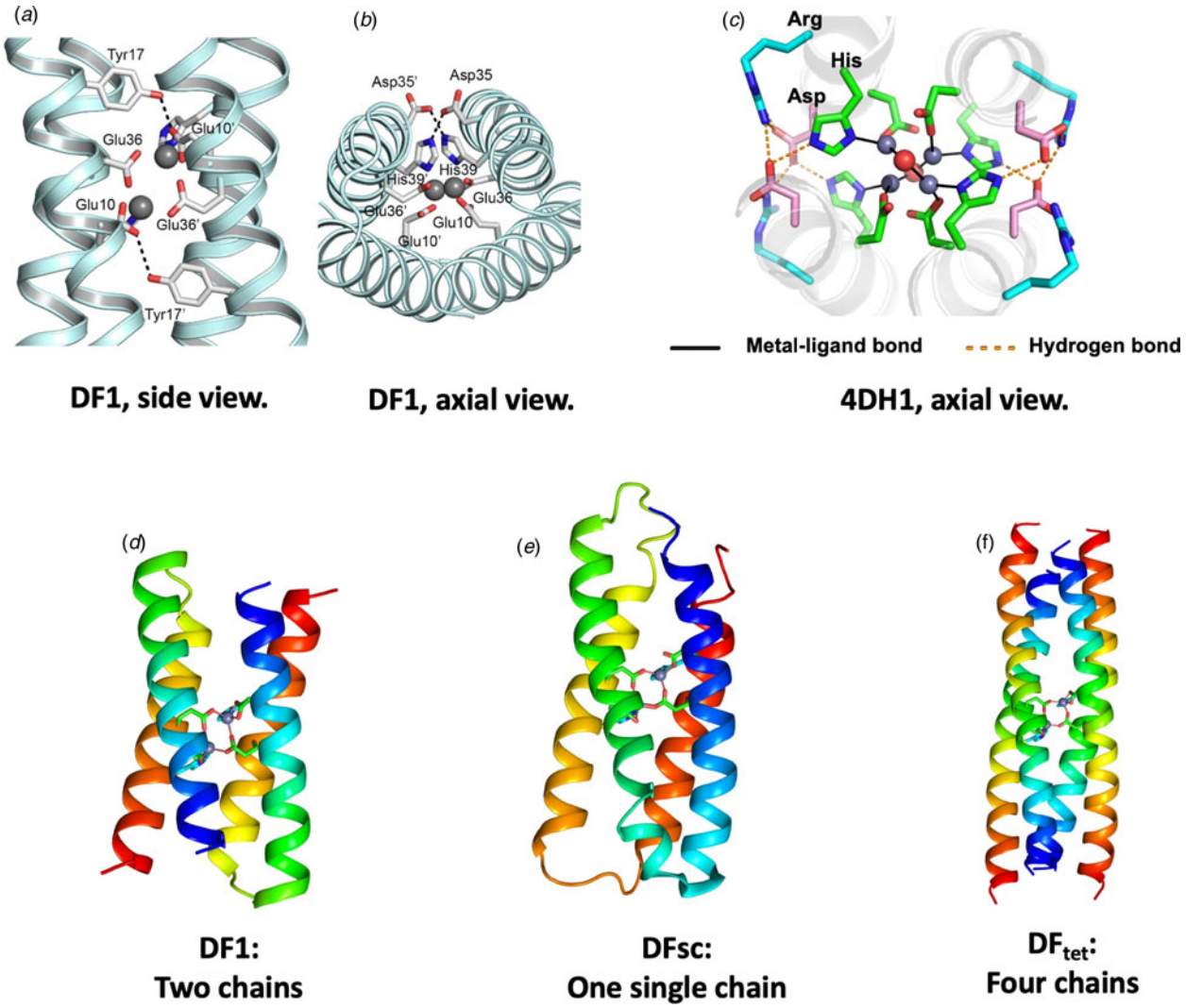


Fig. 6. Design of DF family proteins. Panels (*a–c*) show experimentally determined structures of extended metal-ligand and second-shell hydrogen-bonded networks in DF1 and related proteins. Two projections of DF1s metal-binding site are shown in (*a*) and (*b*) (PDB: 1EC5). Panel (*c*) shows an axial view of 4DH1 (PDB: 5WLL), a DF analog that binds four Zn(II) ions. An Asp residue forms a second-shell-hydrogen bond to a His ligand, and an Arg residue forms a third-shell hydrogen bond. Overall, the network includes four Zn, two waters, eight Asp, four His, and four Arg – all converging at the center of the bundle. Panels (*d–f*) illustrate how the backbone of DF (*d*) was elaborated to create a single chain (DFsc, PDB: 2HZ8) or a self-assembling tetramer (DF_{tet}).

desired changes were highly destabilizing, as they involved burial of additional polar groups (Reig *et al.*, 2012) and removal of Leu sidechains to create an substrate-binding site proximal to the metal ions (DeGrado *et al.*, 2003; Maglio *et al.*, 2003). To compensate, stabilizing substitutions were placed at positions distant from the active site, and an idealized αR-αL-β (Lahr *et al.*, 2005) interhelical loop featuring a network of hydrogen-bonded sidechain/mainchain interactions was installed to favor the folded structure (Faiella *et al.*, 2009). Also, the C_2 symmetry of the initial DF led to functional limitations that could be overcome by building a single-chain version of the protein (DFsc) (Calhoun *et al.*, 2003) with three interhelical loops (Fig. 6*e*).

In an alternate approach, Summa *et al.* designed DF_{tet}, which consisted of four disconnected helices that could be combinatorially assembled to facilitate evaluation of multiple sequence variants for catalytic functions (Fig. 6*f*) (Marsh and DeGrado, 2002; Summa *et al.*, 2002; Kaplan and DeGrado, 2004). To increase stability, the helices of DF_{tet} were extended to 33 residues, and the overall bundle was redesigned to conform to a left-handed coiled coil using an algorithm that incorporates the Crick equations. By engineering the electrostatic interaction at the helix–helix interfaces and an internal hydrogen-bond network, it was possible to design a uniquely folded two-component A_2B_2 tetramer (Summa *et al.*, 2002), as well as a three-component $A_A{\cdot}A_B{\cdot}B_2$ heterotetramer (Marsh and DeGrado, 2002; Kaplan and DeGrado, 2004). Both assembled with very high specificity. A Monte-Carlo algorithm that explicitly evaluated the electrostatic interactions in the desired heterotetramer, as well as other unwanted alternative topologies, facilitated the design. To the best of our knowledge, this was the first use of a computational algorithm to design a sequence that not only stabilized the desired structure (positive design), but also destabilized undesired outcomes (negative design). Since then, sophisticated methods that incorporate machine-learning have been developed for positive and negative design of coiled coils (Grigoryan *et al.*, 2009). Rosetta's H-bond network algorithm can also now facilitate the process of building hydrogen bond networks (Boyken *et al.*, 2016; Chen *et al.*, 2019).

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

A variety of catalytic and binding functions have been engineered into DF protein scaffolds (Lombardi *et al.*, 2019). Precisely as designed, the bespoke site presented unoccupied ligand-binding sites for water, O_2, and organic substrates. By modifying the environment surrounding the diiron site it has been possible to design DF analogs that catalyze the O_2-dependent oxidation of dihydroquinones (Faiella *et al.*, 2009) and amino phenols (Kaplan and DeGrado, 2004) at rates approaching that of the alternative oxidase enzyme. Furthermore, by asymmetrically introducing an additional His ligand (and additional second- and third-shell hydrogen bonding groups) the DF protein has been further engineered to catalyze aniline hydroxylase, mimicking a family of related non-heme enzymes (Reig *et al.*, 2012; Snyder *et al.*, 2015). Finally, a DFsc variant was designed to stabilize the radical semiquinone anion, which is otherwise unstable in aqueous solution (Ulas *et al.*, 2016). The protein stabilized the semiquinone by reducing the midpoint potential for its formation *via* the one-electron oxidation of the catechol by approximately 400 mV (9 kcal mol^{-1}). Hence, the stability of a radical species was drastically stabilized by harnessing its binding energy to the metalloprotein.

Most recently, the design principles used in the construction of DF proteins have recently been extended to engineer tetranuclear Zn^{2+} clusters (Chino *et al.*, 2018; Zhang *et al.*, 2018*a*). The site included four bridging Asp and four terminal His ligands, as well as a total of 16 polar side chains in a fully connected hydrogen-bonded network (Fig. 6*c*). Similar to DF_{tet}, the designed proteins have clusters of apolar sidechains above and below the binding site, which drive the assembly of the bundle. Solution NMR and crystallography confirmed that the desired structure, including a vast network of hydrogen-bonded interactions had indeed been achieved.

Trigonal binding sites in three-helix bundle

Many metal ions are bound in a trigonal geometry, for example, representing three vertices of a tetrahedron, a trigonal pyramid, an octahedron, or a trigonal planar arrangement. The three-helix bundle is particularly compatible with this geometry, and early work with template-assembled peptides using, for example, bi-pyridyl-metal ion interactions (Ghadiri and Case, 1993), achieved this geometry.

In the 1990s, Pecoraro, DeGrado, and coworkers designed the first three-helical bundle metalloproteins, which interacted with Hg(II) in an unusual three-coordinate 3-Cys geometry (Fig. 5) (Dieckmann *et al.*, 1997, 1998). Building on this early success, the Pecoraro lab has greatly expanded the field of *de novo* designed metalloproteins. His group has generated a number of metal complexes that are not known in nature, but can be assembled through *de novo* protein design. The three-fold symmetry of the bundle is ideal for binding metal ions such as Zn (II), Hg(II), Cd(II), Pb(II), As(III), and Bi(III) that prefers lower coordination numbers. The metal binding sites were created by introducing cysteine residues in the '*a*' position of the coiled-coil heptad at various locations in the bundle. The resulting proteins showed mid-nM affinities for cadmium, lead and mercury. Spectroscopic studies, including extended X-ray absorption fine structure (EXAFS), ^{113}Cd, ^{207}Pb, and ^{199}Hg NMR as well ^{113m}Cd and ^{199m}Hg PAC helped elucidate fine structural details of the coordination sphere, which allowed for further fine-tuning of metal coordination sphere (Chakraborty *et al.*, 2010, 2011; Iranzo *et al.*, 2011).

The formation of catalytically competent metal-binding sites in metalloproteins often requires the energetically unfavorable burial of a large number of polar residues in the hydrophobic interior. Pecoraro and coworkers reasoned that the structural stability imparted by the above-mentioned 3-Cys sites might be used to stabilize a second catalytically active metal-binding site within the same bundle. Using this principle they used the 3-Cys Hg(II)-binding site as a structural site to support a second catalytic three-His Zn-binding site. The resulting protein was a remarkably efficient catalyst of CO_2 hydration ($k_{cat}/K_M = 1.8 \times 10^5\ M^{-1}\ s^{-1}$ at pH 9.5) within 500-fold of carbonic anhydrase (Zastrow *et al.*, 2012). The designed zinc-based active site has also been transplanted into α3D (the anti-parallel 73-residue single-chain three-helical bundle protein discussed above) by mutating three of the core leucine residues to histidines (Fig. 3*c*) (Zastrow and Pecoraro, 2013*a*, 2013*b*). The single-chain antiparallel topology is inherently more stable than the self-assembled trimeric bundle; therefore the 3-Cys structural site is no longer necessary. The resulting metalloenzyme $Zn^{II}\alpha$3DH3 efficiently promotes *p*-NPA hydrolysis and CO_2 hydration. Its kinetic parameters are somewhat lower than those of the 3-chain predecessor; however, due to its single-chain topology, it can be improved using directed evolution.

To expand the repertoire of catalyzed chemical reactions by *de novo* designed trimeric coil coils enzymes to redox transformations, Pecoraro explored copper binding of TRIL23H, a close relative of metallohydrolase-supporting peptide TRIL9CL23H, but without the mercury structural site. TRIL23H binds Cu(II) with nM–μM affinity and Cu(I) with pM affinity fulfilling the key requirement for redox cycling. The copper ion in Cu(I/II) $(TRIL23H)_3$ is bound by three histidine residues leaving two sites open to substrate/reductant coordination in a manner similar to that of the Cu_{T2} center of copper nitrite reductase (Tegoni *et al.*, 2012). The designed metalloenzyme catalyzes reduction of nitrate to NO using ascorbate as the ultimate reductant for at least five turnovers.

The functional versatility of the trimeric coiled coils goes beyond catalysis. Peacock and co-workers have successfully utilized them to create magnetic resonance imaging probes with excellent relaxivity properties (Berwick *et al.*, 2014, 2016). Tanaka and coworkers extended the trimeric helical Ile zipper peptides described by Alber and coworkers to create a 3-His site capable of binding transition metals with different geometries (Suzuki *et al.*, 1998; Kiyokawa *et al.*, 2004; Tanaka *et al.*, 2004). The ability of the resulting peptides to oligomerize in a predicable manner was used to induce trimerization of DNA-binding domains of the heat shock proteins from *Saccharomyces cerevisiae* (Murase *et al.*, 2012). Fusing a variant of the green fluorescent protein to metal-binding coiled coils produced fluorescent sensors for metal ions (Murase *et al.*, 2012).

Directed evolution of the esterase activity of a Zn^{2+}-binding helical bundle built on a natural protein scaffold

In the process of creating a metal-mediated protein–protein interface, Kuhlman and co-workers discovered MID1, a zinc-binding dimeric helix–loop–helix protein that can promote *p*-nitrophenol ester and phosphoester hydrolysis with reasonable catalytic efficiencies (Der *et al.*, 2012*b*). While this work involved modification of an existing natural protein rather than full *de novo* design, the fold used was of similar complexity to the *de novo* scaffolds discussed above, allowing comparison of the two approaches. The Rosetta Match algorithm was used to identify protein structures

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

from the PDB that could form half of a tetrahedral Zn^{2+} binding site when His or Cys ligands were introduced at appropriate surface locations. In the design strategy, a complete tetrahedral site was formed when the proteins associated to form symmetrical homodimers. A total of 600 natural protein scaffolds were screened, resulting in 1.5 million design trajectories, which were evaluated over 25 000 cpu hours. Eight designs were experimentally evaluated, and one, designated MID1, was sufficiently well behaved to allow characterization. In the intended design, MID1 contains two symmetrically related Zn^{2+}-consisting of His residues at positions i and $i+4$ introduced along the surface of a small helix–loop–helix domain from rabenosyn. This arrangement had been used for many years to mediate Zn^{2+} binding in *de novo* designed peptides (Ghadiri and Choi, 1990; Ruan *et al.*, 1990; Krantz and Sosnick, 2001; Tang *et al.*, 2007; Signarvic and DeGrado, 2009) as well as Zn^{2+}-mediated dimerization of *de novo* designed proteins (Handel and DeGrado, 1990; Handel *et al.*, 1993) and natural proteins (Salgado *et al.*, 2007, 2010).

NMR and crystallographic analysis of MID1 showed considerable plasticity, with both similarities as well as differences to the design. As in the design each i, $i+4$ His residue ligated a single ion via the ϵ-nitrogen. However, the third His bound in an unexpected geometry via the δ nitrogen, and the fourth His did not ligate Zn^{2+} at all (Der *et al.*, 2012*b*).

These differences were surprising given the above-mentioned successes in *de novo* metalloprotein design, in which the functional requirements were used to define both the fold and the site. Moreover, small perturbations to MID1, such as single-site amino acid substitutions or changing the metal ion from Zn^{2+} to Co^{2+} caused large changes in the helix-packing geometry of MID1 (Fig. 7) (Der *et al.*, 2012*b*). Serendipitously, MID1 had a weak 4-nitrophenyl esterase activity associated with the unexpected 3-His binding geometry, which resulted in a free ligation site on the bound Zn^{2+} (Der *et al.*, 2012*a*).

Similarly, Song and Tezcan introduced zinc binding sites into cytochrome bc_{562} to promote controlled self-assembly into tetrameric species. The resulting assembly promotes hydrolysis of various substrates (Song and Tezcan, 2014).

The plasticity of the MID1 protein proved beneficial for *in vitro* evolution of a stereoselective metalloenzyme capable of hydrolysis of model fluorogenic substrates (Studer *et al.*, 2018). A single-chain version of MID1, $MID1_{sc}$ with a single metal-binding site served as the starting point for *in vitro* evolution. In all five rounds of cassette mutagenesis, two rounds of random mutagenesis, and two rounds of DNA shuffling were employed. Ultimately, a catalytic efficiency of $k_{cat}/K_M = 980\,000\ M^{-1}\ s^{-1}$ ($k_{cat} = 1.6\ s^{-1}$; $K_M = 1.6\ \mu M$) was achieved, highlighting the power of directed evolution in combination with rational protein design (Studer *et al.*, 2018). The crystallographic structure of the resulting protein, $MID1_{sc}10$ showed that the protein had undergone a number of remarkable changes in the course of evolution. One of the His ligands was lost and another gained at a different location, resulting in a 7 Å translation of the metal-binding site. Moreover, a substrate-binding site was created by multiple substitutions as well as a large, rigid-body rotation of one helix–turn–helix motif (Fig. 7).

It is instructive to compare the contributions of the metal ion *versus* the protein to the esterase activity of $MID1_{sc}$ *versus* some of the purposefully designed proteins discussed above. The value of k_{cat}/k_{uncat} for $MID1_{sc}$ is 1.6×10^5, while that of a designed amyloid-forming Zn^{2+}-binding heptapeptide IHIHIQI is 100-fold lower (1.6×10^3) at the same pH (Rufo *et al.*, 2014). The heptapeptide has a similar 3-His active site capable of activating a water molecule for hydrolysis (Lee *et al.*, 2017), but lacks cavities to bind the substrates. By contrast, $MID1_{sc}$ has a deep pocket capable of stereospecific binding of the large hydrophobic substrate, **1** (Fig. 7) used in the directed evolution experiments. The substrate-binding interactions result in considerable stereospecificity for **1** and a relatively tight K_M of 1.6 μM. By comparison, both $MID1_{sc}10$ and IHIHIQI hydrolyze the minimal substrate, **2**, with similar values of k_{cat}/K_M ($32\ M^{-1}\ s^{-1}$ for MID1 *versus* $62\ M^{-1}\ s^{-1}$ for IHIHIQI), likely reflecting the contribution of the preorganized metal complex. The additional catalytic efficiency of $MID1_{sc}10$ for substrate **1** likely reflects more precise positioning of the substrate for attack in the Michaelis complex. These studies show the power of directed evolution to create substrate-binding interactions that work in concert with a metal to produce significant rate enhancements.

Helical bundles as catalysts and inhibitors of protein–protein interactions

Four-helix bundles were also used to test concepts of catalysis and to design inhibitors of protein–protein interactions. Baltzer and co-workers employed this strategy to design catalytic proteins. A 42-residue peptide KO-42 assembles into an antiparallel four-helix bundle with catalytic sites engineered on the surface of the bundle as demonstrated by NMR, circular dichroism (CD) spectroscopy and ultracentrifugation, to catalyze hydrolysis of *p*-nitrophenyl esters with a rate enhancement of three orders of magnitude compared to the imidazole control (Broo *et al.*, 1997). Subsequent rational improvement of the design allowed for introduction of enantioselective recognition of substrates (Broo *et al.*, 1998), a hallmark of natural proteins, and for elucidation of the role the pK_a of the active residue as well as the geometry of the active site on catalysis (Broo *et al.*, 1998; Baltzer *et al.*, 1999). Expansion of the active site in the bundles to include additional residues to provide transition state stabilization allowed for hydrolysis of challenging phosphoester substrates, including uridine 3′−2,2,2-trichloroethylphosphate, a mimic of RNA (Razkin *et al.*, 2007, 2008). The simple architecture of KO-42 is nonetheless amenable to introduction of binding sites for complex substrates, whose recognition relies on multiple substrate–protein interactions. In addition to a histidine-based active site to promote proton-transfer, KO-42 was modified to incorporate positively charged residues to stabilize negatively charged aldimine. The resulting peptide bundles T-4 and T-16 promote aldimine to ketamine conversion, emulating biosynthetic transamination reactions (Allert and Baltzer, 2003). Finally, the graded reactivity of KO-42 has been used to allow the site-directed assembly of auxiliary binding groups, to create binders of protein surfaces with sub-nanomolar affinity for the proteins of interest (Baltzer, 2011; Yang *et al.*, 2017*a*, 2017*b*).

While the binding and catalytic sites of derivatives of KO-42 lie along the surface of the bundle, Woolfson and coworkers used the hollow surface of *de novo* designed proteins to create functional sites. They succeeded in building a catalytic dyad in a peptide that self-assembles into a heptameric coiled coil with no known natural analogs that promotes ester hydrolysis (Burton *et al.*, 2016).

Helical bundles have been designed or selected to bind to a variety of other protein surfaces, to create inhibitors of protein–protein interactions (Fujiwara and Fujii, 2013; Fujiwara *et al.*, 2016). A recent example illustrates how far *de novo* protein design has progressed from the early days of parametric helical bundle

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

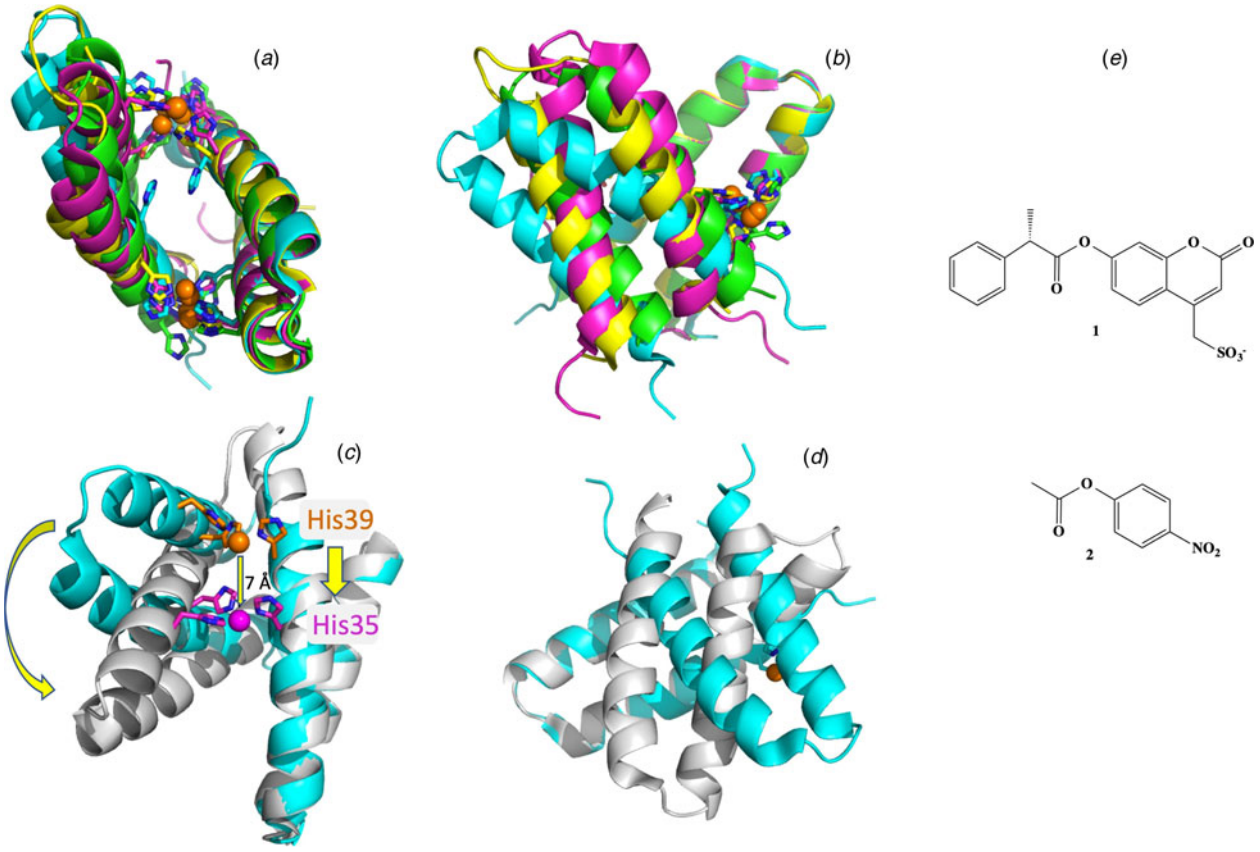


Fig. 7. Structural plasticity of MID1 (*a* and *b*). Two views of the crystal structures of di-zinc MID1 (PDB: 3V1C, blue ribbon), di-cobalt MID1 (PDB: 3V1D, magenta), di-zinc MID1-H12E (PDB: 3V1E, yellow), and di-zinc MID1-H35E (PDB: 3V1F, green) are shown with one of the two helix–loop–helix motifs superimposed. The overlay shows the variability in metal ion positions and ligand geometry, as well as variations in inter-subunit interactions. Panels (*c*) and (*d*) illustrate a similar superposition of di-zinc MID1 (PDB: 3V1C, blue ribbon, orange carbon atoms as sticks) with di-Zinc $MID1_{sc}10$ (PDB: 5OD1, gray ribbon, magenta C atoms as sticks) showing a large rigid-body rotation of the helical hairpins, a shift in the primary ligand from His39 to His35, and a 7 Å shift of the metal ion. Panel (*e*) shows the substrates used to characterize the catalytic activity of $MID1_{sc}10$.

design of proteins to incorporate the sophisticated computational design algorithms in Rosetta as well as directed evolution and sequence display of combinatorial libraries in the work flow. Baker and coworkers recently combined these technologies to design mimics of interleukin-2 (IL-2) that bind to the IL-2 receptor $\beta\gamma$c heterodimer (IL-2R$\beta\gamma$c), but not to IL-2Rα or IL-15Rα. The designs used the natural four-helix bundle, IL-2, as a starting point. In a series of steps the IL-2 bundle was progressively idealized using parametric protein design, and its folding topology was simplified by introduction of short idealized loops. At each round of design, the sequences were experimentally evaluated and the affinity was enhanced by multiple rounds of display on yeast. Crystal structures of an optimized design protein alone and in complex with IL-2R$\beta\gamma$c, are very similar to the designed model. The family of designed proteins has superior therapeutic activity to IL-2 in mouse models of melanoma and colon cancer, with reduced toxicity and undetectable immunogenicity.

Helical bundles for binding complex cofactors

Dutton and DeGrado utilized a sequence-based approach to design heme-binding proteins designated 'maquettes' to probe the function of multi-heme proteins. A 31-residue long peptide designed to mimic the key structural features of cytochrome bc_1 was shown to assemble in the presence of four hemin moieties to form a four-helix bundle. Introduction of a flexible Cys containing linker allowed for further stabilization of the structure, effectively creating a helix–loop–helix motif (Robertson *et al.*, 1994). The original designs have been elaborated by Dutton, Moser, Gibney, Anderson, and coworkers to include complex single-chain topologies that allowed sequence diversification and recombinant expression (Grayson and Anderson, 2018). The simple geometry of maquettes (Fig. 8*a*) allowed for direct elucidation of factors that define electrochemical properties of heme in metalloproteins and subsequent rational tuning of the redox potential of the cofactors (Kennedy and Gibney, 2001; Reedy and Gibney, 2004). Subsequent studies show that the maquette architecture can support diverse protein functionalities ranging from light capture to catalysis (Koder *et al.*, 2009; Lichtenstein *et al.*, 2012; Kodali *et al.*, 2017; Watkins *et al.*, 2017). The malleable, dynamic maquette scaffolds bind cofactors with high affinity and serve as starting points for further improvement supporting the notion that substantial initial level of functionality is fairly easy to achieve in *de novo* designed proteins. Nevertheless, it did not prove to be possible to solve solution NMR or crystallographic structures of the family of maquettes with their cofactors bound. One structure was solved for an apo-structure, but the structure was not compatible with the requirements of binding heme (Huang *et al.*, 2004).

Multiheme-binding helical bundles can also be designed completely *de novo* based on parameterized backbones, the first being closely related to α4 (Choma *et al.*, 1994). Subsequent parameterizations were based on positioning keystone residues for first- and second-shell ligation as well as steric packing. This

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

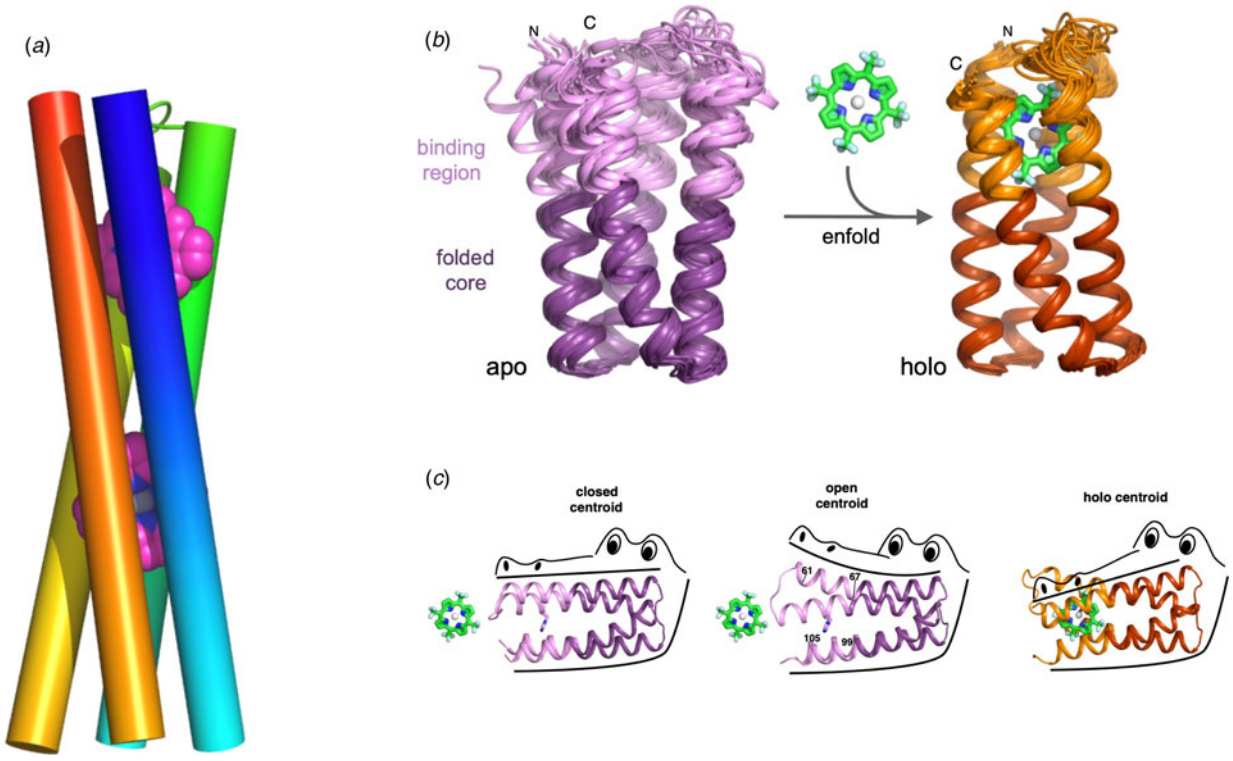


Fig. 8. Cofactor-binding helical bundles. Panel (*a*) shows a model of a two-porphyrin maquette. High-resolution structures have not been published for cofactor-bound maquettes, likely due to dynamic properties (Koder *et al*., 2009; Lichtenstein *et al*., 2012; Kodali *et al*., 2017; Watkins *et al*., 2017). However, recent work on other *de novo* proteins including PS1 indicates that it is possible to design uniquely structured porphyrin-binding proteins (Polizzi *et al*., 2017). Panels (*b*) and (*c*) illustrate PS1, a porphyrin-binding protein, that was instead computationally designed to carefully optimize the packing of the core as well as the packing of the cofactor (Polizzi *et al*., 2017). The high-resolution solution structure of the apo-state has two conformations that appear to facilitate binding of the porphyrin. Both conformers have well-packed hydrophobic core, but differ in the orientation of the helices in the binding site. Binding of the porphyrin results in ordering of the entire protein.

approach was expanded to enable design of a variety of cofactors that contain various metals (Bender *et al*., 2007; Fry *et al*., 2010, 2013; Korendovych *et al*., 2010).

Only recently has the successful design of a porphyrin-binding protein with sub-Ångstrom accuracy been accomplished as verified by high-resolution structure determination. The key was to consider what had traditionally been considered as separate sectors – the hydrophobic core and ligand-binding site – inseparable units (Figs 8*b* and *c*) (Polizzi *et al*., 2017). Flexible backbone design of a parametrically defined protein template allows to simultaneously pack both the protein interior both proximal to and remote from the ligand-binding site. Thus, tight interdigitation of core side chains quite removed from the binding site structurally can cooperate to restrain and stabilize the first- and second-shell packing around the ligand. The resulting protein, PS1, bound an electron-deficient, non-natural porphyrin at temperatures up to 100 °C, and its structure was in sub-Ångstrom agreement with the design. These results illustrated the unification of core packing and binding site definition as a central principle of ligand-binding protein design. It also bodes well for the design of 'maquettes' that are uniquely structured, rather than multi-conformational in nature.

Beyond helical bundles

By 2000, the accurate *de novo* design of homo-oligomeric coiled coils (Harbury *et al*., 1995, 1998; Ogihara *et al*., 1997) and helical bundles such as α3D and DF (Lombardi *et al*., 2000*b*) had been accomplished. By contrast, the design and structure determination of uniquely folded globular proteins containing β-structure remained problematic. Early attempts to design an all-β protein called betabellin resulted in structures with poor solubility (Richardson and Richardson, 1989), likely due to amyloid formation (Lim *et al*., 1998). Analysis of the failures, however, led to important insights (Richardson and Richardson, 2002). The edges of β-sheets are sticky sites that can engage in aggregation and amyloid formation. In natural proteins, such aggregation is minimized by decreasing the length of edge strands and endowing them with Pro residues or polar groups that decrease interchain hydrogen-bonding and hydrophobic interactions that can lead to oligomerization.

Nevertheless, in the 1990s significant progress was made toward the design of peptides that form β-hairpins, including the Trp zipper peptides that displayed well-defined β-hairpin conformations stabilized by cross-strand pairs of indole rings (Cochran *et al*., 2001). Also, by 1998, several groups had demonstrated the design of three-stranded β-sheets, with varying degrees of water-solubility and stability (Das *et al*., 1998; Kortemme *et al*., 1998; Schenck and Gellman, 1998; Sharman and Searle, 1998). As mentioned above, Dahiyat and Mayo had also succeeded in the fully automated redesign of the sequence of a zinc finger peptide, resulting in a peptide that folded into a structure consisting of an α-helix packed against an antiparallel β-hairpin (Richardson and Richardson, 2002). Imperiali's group also redesigned a similar zinc finger to produce a peptide that folded in the absence of metal ions (Struthers *et al*., 1996*a*, 1996*b*).

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

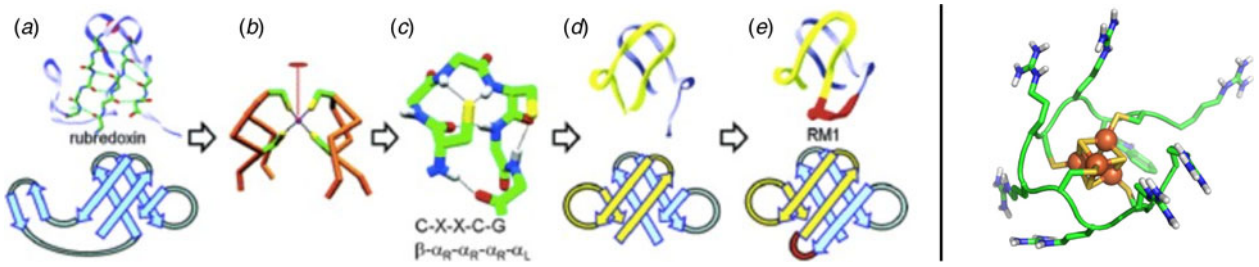


Fig. 9. (Left) RM1 design cycle: (*a*) three-stranded sheet topology of natural rubredoxin, (*b*) C_2 symmetry, (*c*) active-site geometry, (*d*) miniRM dimer, and (*e*) RM1 with Trpzip linker shown in red. Reproduced with permission from Nanda *et al*. (2005). Copyright (2005) American Chemical Society. (Right) Computational model of ambidoxin.

The fundamental parameterization approach described above is by no means limited to helical bundles. Lombardi and coworkers built METP, a β-hairpin miniaturized electron transfer protein, by parameterizing the metal-binding site of a natural rubredoxin (Lombardi *et al*., 2000*a*). Nanda, DeGrado and coworkers designed RM1, a stable minimalist protein that folds in a β-sheet structure both in the presence and in the absence of iron (Fig. 9). RM1s design was based on a simple dimeric sheet–turn–sheet secondary motif. RM1 binds iron to form a stable, redox-active four-cysteine thiolate iron site that is structurally and functionally analogous to that of rubredoxin (Nanda *et al*., 2005). Recently, Nanda, Fialkowski, and coworkers have been able to design ambidoxin, a 12-residue peptide with alternating D- and L-amino acid residues to stabilize a functional 4Fe–4S cubane cluster through metal–side chain interactions and an intricate network of hydrogen bonds (Kang *et al*., 2018).

Membrane protein design

There are two structural classes of TM proteins: β-barrels that are found in the outer membranes of bacteria and mitochondria and the helical bundles, which are found in cytoplasmic and organelle membranes. Given the greater functional diversity of the helical bundle class of membrane proteins, most work in *de novo* design has focused on this class of membrane proteins. *De novo* membrane protein design has contributed significantly to understanding fundamental principles by which membrane proteins achieve their folded conformations and functions such as active ion transport.

Understanding the rules of membrane protein folding, stability, and assembly

Helical membrane proteins begin folding as they exit the translocon, completing the process in the membrane environment (Engelman *et al*., 2003; White and von Heijne, 2008). The folding of membrane proteins thus can be minimally approximated by a two-stage process involving the biosynthetic or physical insertion of TM helices into membranes followed by their subsequent assembly to form native structures. The features required for insertion are well understood from elegant studies of von Heijne, White, and others who examined the sequence-dependence of helix insertion into membranes via the translocon (Hessa *et al*., 2005; White and von Heijne, 2005, 2008). The resulting 'biological hydrophobicity scale' was in good agreement with those obtained from model compounds as well as scales derived from structural informatics of membrane proteins (Senes *et al*., 2007; Schramm *et al*., 2012). Such information has long provided restraints for design of monomeric helical peptides that insert into membranes (Ren *et al*., 1999; Morein *et al*., 2000; Caputo and London, 2003), and has been incorporated into programs for membrane protein design, such as Rosetta Membrane (Elazar *et al*., 2016; Koehler Leman *et al*., 2017; Duran and Meiler, 2018).

De novo design has contributed to understanding the next key step in membrane protein folding when helices laterally associate to form an inter- or intra-molecular TM bundle. Much of the work has focused on engineering assemblies of TM α-helices from single-spanning membrane proteins, chosen for their biological relevance and technical advantages. Over 50% of all membrane proteins are single-spanning, yet they are the least structurally characterized class of MPs. Their lateral TM helix interactions play vital roles in signaling, complex formation, and ion conduction (Kirrbach *et al*., 2013; Lomize *et al*., 2017). Aberrant folding or assembly is also involved in devastating diseases from cancer to Alzheimer's disease (Partridge *et al*., 2004; Schlebach and Sanders, 2015). Additionally, unlike complex multi-pass proteins, single-span TM bundles allow investigation of inter-helical interactions with a clear unfolded state – a monomeric α-helix – free of extracellular domains or loops that cloud interpretation. Moreover, conformational specificity and folding can be simply evaluated by determining whether a single oligomeric state is formed.

Small residue motifs that stabilize TM helix–helix-packing interactions

Some of the earliest studies on TM helix–helix interactions focused on the identification of sequence motifs, such as the GX_3G, found in glycophorin A. GX_3G, or more generally the Small-X_3-Small (in which Small is Gly, Ala, or Ser) motif is involved in both intramolecular folding as well as intermolecular assembly of TM helices (Langosch *et al*., 1996; Brosig and Langosch, 1998; Senes *et al*., 2000, 2001). The small residues line along one face of the helix and mediate a very close approach of the backbones of two helices, which interact with a right-handed crossing angle of near 40° (MacKenzie *et al*., 1997). The interface is stabilized through extensive vdW interactions (Duong *et al*., 2007; Mueller *et al*., 2014) and CH hydrogen bonds between the backbone Cα–H and the carbonyl oxygen of neighboring helices (Senes *et al*., 2001; Arbely and Arkin, 2004; Mueller *et al*., 2014). The stability of the Small-X_3-Small motif depends critically on the position in the membrane as well as the sequence context surrounding the two small residues (Duong *et al*., 2007; Unterreitmeier *et al*., 2007; MacKenzie and Fleming, 2008; Langosch and Arkin, 2009).

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

A second motif that has been used extensively in membrane protein design is an antiparallel zipper-like packing with a Gly, Ala, or Ser in a $(\text{Small-X}_6)_n$ motif (Adamian and Liang, 2002; Walters and DeGrado, 2006). This sequence motif specifies folding into a structure similar to the alanine-coil seen water-soluble proteins (Gernert *et al.*, 1995), with a left-handed crossing angle near −10° to −20°. The presence of a single small residue per heptad enables intimate packing. Computational design of model TM coiled-coil peptides (designated MS1 peptides) with various residues at the '*a*' position showed association strengths in the order: Gly > Ala > Val > Ile. Moreover, MS1-Gly has a strong tendency to form antiparallel dimers, MS1-Ala formed a mixture of parallel and antiparallel dimers, while MS1-Val and MS1-Ile have a preference to form very weakly associating parallel dimers. Calculations based on exhaustive conformational searching and rotamer optimization were in excellent agreement with experiments, in terms of the overall stability of the structures and the preference for parallel *versus* antiparallel packing. These studies demonstrated that vdW interactions and electrostatic interactions contribute to the stability and topological preferences of the dimers.

Hydrogen-bonded interactions can stabilize membrane proteins

Hydrogen-bonds between polar sidechains can also contribute to the stability of membrane proteins. The introduction of strongly polar residues, including Asp, Asn, Glu, and Gln can lead to association of designed TM peptides (Choma *et al.*, 2000; Zhou *et al.*, 2000, 2001; Gratkowski *et al.*, 2001). The energetics of the interaction depends on environment, ranging from very stabilizing near the middle of the TM helix (−2.0 kcal mol^{-1} per Asn side chain) to very weak (0 ± 0.5 kcal mol^{-1}) near the ends of the helix, which locate to the headgroup region (Lear *et al.*, 2003). These data are consistent with the expectation that sidechain hydrogen bonding will contribute to stability in the relatively dry region of a membrane, but not in regions where water can compete for hydrogen bonds in the monomeric state.

Both the thermodynamics and geometric specificity of association of TM helices can be enhanced through the design of an extensive hydrogen-bonded network, as shown in work in which three Asn and three Thr sidechains were engineered to interact in a three-helix bundle (Tatko *et al.*, 2006). More recently, Baker and coworkers have designed elongated membrane-spanning helical bundles, which contain hydrogen-bonded networks built using the HB-net module of Rosetta (Lu *et al.*, 2018). The designs included chains with two TM helices, representing the first examples of the *de novo* design of multi-pass membrane proteins, whose crystallographic structures were determined at high resolution.

Contribution of packing of large apolar residues to the stability of membrane proteins

All of the above designed membrane proteins relied on either polar interactions to drive assembly in the membrane, or small residues at appropriate spacings to drive folding through close contacts between the backbones of helices. However, such motifs, although not uncommon, are not a general feature of the interhelical packing seen in natural membrane proteins. Instead, helix–helix packings are stabilized by interactions of apolar sidechains, similar to that in water-soluble proteins. The hydrophobic effect, which represents the predominant driving force for protein folding in water, is negligible in lipid membranes. Thus, in membrane proteins it was unclear whether analogous side-chain packing in the native state provides significant structural stabilization. On the one hand, the same apolar moieties pack similarly with lipid tails in the exposed unfolded state. On the other hand, side-chains pack slightly more efficiently in membrane proteins (Eilers *et al.*, 2000; Adamian and Liang, 2001; Oberai *et al.*, 2009; Zhang *et al.*, 2015), and hence might stabilize folding via favorable vdW interactions and possibly also lipid-specific effects such as 'solvophobic' exclusion (Langosch and Heringa, 1998; Joh *et al.*, 2009; Hong, 2014; Anderson *et al.*, 2017). Mutations to membrane proteins that strongly disrupt vdW packing in the protein interior, either by introducing voids or steric clashes, have been shown to destabilize their native state to various degrees (Doura *et al.*, 2004; Joh *et al.*, 2009; Baker and Urban, 2012; Guo *et al.*, 2016). Nevertheless, it was less clear whether apolar packing can play a dominant role in membrane protein folding, or whether this feature is secondary to other more stabilizing interactions discussed above such as hydrogen bonding and weakly polar C–H-hydrogen bonds. If apolar packing contributed largely to the stabilization of membrane proteins it should be possible to design them based on this feature alone. However, for a number of years this proved to be very difficult (Whitley *et al.*, 1994; Gurezka *et al.*, 1999; Choma *et al.*, 2000; Zhou *et al.*, 2001; Yano *et al.*, 2002; Johnson *et al.*, 2004).

Building on the design principles discovered by Woolfson *et al.*, in the construction of multistranded water-soluble coiled coils, Mravic *et al.* recently designed a homo-pentameric TM five-helix bundle stabilized by apolar packing in the membrane-spanning region alone (Mravic *et al.*, 2019). Successful design required consideration of not only the '*a*' and '*d*' residues in the core, but also the more interfacial '*e*' and '*g*' residues. The resulting pentamers were remarkably stable, even at boiling temperatures in sodium dodecylsulfate. In spite of this extraordinary stability, the steric complementarity required for their folding was shown to be remarkably stringent when compared to helix–helix packings of water-soluble proteins. Thus, substitutions of Leu to Ile entirely disrupted any association of the helices. A strong hydrophobic driving force dominates folding in water, so natural proteins need not achieve stringent packing to fold. Without a hydrophobic force in bilayers, it appears geometric complementarity must be more strictly optimized to achieve folding in membrane proteins. Structural informatics shows that the designed packing motif recurs across the TM proteome, emphasizing a significant role for precise apolar packing in membrane protein folding and stabilization.

Design of functional membrane proteins

Design of TM proteins capable of proton, metal ion, and electron transfer

Given recent progress in designing membrane proteins with predetermined structures, it should be increasingly possible to design function as well. In fact, there have already been some significant accomplishments in the *de novo* design of proteins capable of transporting protons, ions, and electrons. The first TM helical bundles were designed in the late 1980s as functional models for proton channels and ion channels – significantly before the first high-resolution structures of proteins of this class had been determined (Lear *et al.*, 1988; DeGrado *et al.*, 1989). To gain insight into the mechanisms by which α-helices in channels associate and conduct ions, several peptides containing only Leu and Ser residues were designed and computationally modeled. A 21-residue peptide $(\text{Leu-Ser-Ser-Leu-Leu-Ser-Leu})_3$, formed well-

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

defined ion channels with single-channel conductance characteristics resembling the acetylcholine receptor. A second peptide $(Leu\text{-}Ser\text{-}Leu\text{-}Leu\text{-}Leu\text{-}Ser\text{-}Leu)_3$, in which one Ser per heptad repeat was replaced by Leu, produced proton-selective channels. Computer graphics and energy minimization were used to create molecular models that were consistent with the observed properties of the channels. The deduced structures were helical bundles with left-handed crossing angles between the helices. The packing of small Ser residues in a zipper-like manner dictated a tetrameric arrangement for the proton channel $(Leu\text{-}Ser\text{-}Leu\text{-}Leu\text{-}Leu\text{-}Ser\text{-}Leu)_3$ (Fig. 10*a*) and a hexameric channel for the ion channel forming $(Leu\text{-}Ser\text{-}Ser\text{-}Leu\text{-}Leu\text{-}Ser\text{-}Leu)_3$. The hydroxyl sidechains of the Ser residues interacted with water to create a pore large enough to accommodate a solvated ion in the hexameric bundle. The tetrameric bundle was more tightly packed but had voids large enough to accommodate water molecules that appeared to form a proton conduction pathway via a water-hopping mechanism. While crystallographic structures were not available at the time, a large body of subsequent data supported the underlying hypothetical structures and conduction model (DeGrado *et al.*, 1989; Åkerfeldt *et al.*, 1992, 1993; Zhong *et al.*, 1998; Dieckmann *et al.*, 1999; Randa *et al.*, 1999; Nguyen *et al.*, 2013).

The most ambitious functional membrane protein designed to date is a TM four-helix bundle, Rocker (Fig. 10*c*), that transports first-row transition metal ions Zn^{2+} in exchange for protons (Joh *et al.*, 2014, 2017). The design of a Zn^{2+}/proton transporter presented several grand challenges: the first was the design of a membrane protein with a predetermined structure, and the determination of its structure and dynamics at high resolution (which had not yet been accomplished for a *de novo* membrane protein). Next, the design should precisely position polar ionizable Zn^{2+} ligands, which ordinarily are excluded from a membrane environment. Furthermore, to achieve antiporting, it was important to thermodynamically link the binding of protons to changes in the affinity for metal ions. Finally, it was important to anticipate and orchestrate dynamics to facilitate transport of an ion through the channel.

Joh, Grigoryan, and DeGrado designed Rocker using four helices that present metal-binding sites similar to those used in the water-soluble DF proteins discussed above. Previously, Pasternak *et al.* had found that the Glu sidechains in a 4Glu-2His di-Zn^{2+}-binding DF protein were largely protonated in the metal-free apo state, due to the energetic cost of burying negatively-charged sidechains within the interior of a protein (Pasternak *et al.*, 2001). Binding of Zn^{2+} displaces these protons, providing a means to achieve the desired thermodynamic coupling. A computational design algorithm was next used to stabilize two energetically degenerate asymmetric states of the protein while destabilizing a competing fully symmetrical state which might otherwise bind metal ions too tightly and impede motions required for ion transport. The computed TM bundle formed a dimer of dimers with two non-equivalent helix–helix interfaces (Fig. 10*c*); a 'tight interface' had a small inter-helical distance (8.9 Å) stabilized by efficient packing of small, Ala residues. The 'loose interface' had a larger interhelical distance of 12.0 Å and was less well packed. The resulting membrane-spanning four-helical bundle transported first-row transition metal ions Zn^{2+} and Co^{2+}, but not Ca^{2+} across membranes. X-ray crystallography and solid-state and solution NMR confirmed that the overall helical bundle was composed of two tightly interacting pairs of helices, which interacted along the more dynamic interface. Vesicle flux experiments show that as Zn^{2+} ions diffuse down their concentration gradients, protons were antiported. These experiments illustrate the feasibility of designing membrane proteins with predefined structural and dynamic properties.

TM electron transfer is a critical part of the bioenergetic processes that power life. Electrons are transmitted across membranes by hopping between redox-active cofactors. Discher, Dutton, and co-workers have utilized the maquette scaffold to create helical bundles that contain both soluble and TM domains for light harvesting and electron transfer (Ye *et al.*, 2005; Goparaju *et al.*, 2016). Korendovych and coworkers designed a TM four-helix bundle PRIME that bound two iron-porphyrin cofactors in a bis-His geometry. The resulting protein is perfectly suited to catalyze the transfer of electrons across phospholipid bilayers (Fig. 10*b*) (Korendovych *et al.*, 2010). Analytical ultracentrifugation, EPR, redox potentiometry and UV-visible CD spectroscopy showed that the desired complex had been formed. Moreover, the protein bound the targeted di-phenyl-porphyrin derivative with high affinity and high specificity relative to other porphyrin or heme derivatives. Thus, both cofactor binding and TM electron transfer were realized for the first time in a *de novo* TM bundle.

De novo *design of TM peptides that recognize the TM helices of natural proteins*

While there are a large number of reagents such as antibodies that are capable of recognizing water-soluble proteins or the extramembrane regions of membrane proteins, there is a great need to develop equivalent reagents to target the membrane-spanning regions of TM proteins. Such reagents could be used to interrogate the interactions between TM helices in natural proteins. The Small-X_3-Small motif has been used to design peptides that specifically recognize the TM domains of two different integrins. Integrins are heterodimers with single-TM helices that tightly interact in the resting state, but separate in the activated state. Yin *et al.* achieved the computation design of peptides that specifically recognize the TM helices of two closely related integrins ($\alpha_{IIb}\beta_3$ and $\alpha_V\beta_3$) in micelles, bacterial membranes, and mammalian cells (Yin *et al.*, 2007; Caputo *et al.*, 2008). The peptides competed for the endogenous helix–helix interactions and hence activated the integrins in a sequence-specific manner. These data showed that sequence-specific recognition of helices in TM proteins can be achieved through optimization of the geometric complementarity of the target-host complex. Less sequence specificity was observed in more recently designed peptides that target $\beta1$ integrins. Nevertheless, very useful reagents were obtained to target and activate this class of proteins (Mravic *et al.*, 2018).

Fragment-based and bioinformatically informed computational protein design

Backbone fragments and sequence statistics broaden the scope of protein design

Despite the success described above in the sections on parametric design of water-soluble proteins, the *de novo* design of larger cooperatively folded proteins rich in β-sheets remained problematic until recently (Hecht, 1994; Quinn et al., 1994; Yan and Erickson, 1994). It was therefore of great interest when Kuhlman and Baker described the design of TOP7 (Fig. 11*a*), a

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

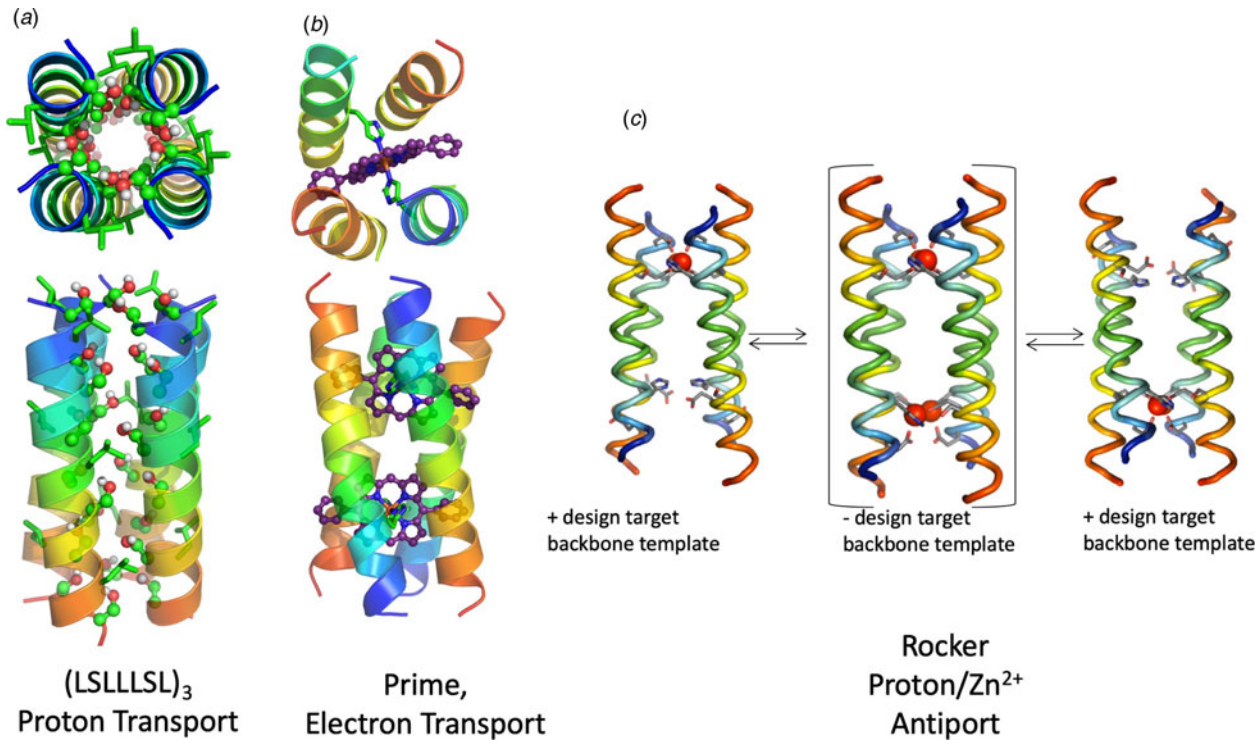


Fig. 10. (*a*, *b*) Top and side views of computational models of *de novo* designed ion pores LS2 and PRIME, respectively. In panel (*a*), the Ser sidechains of LS2 are shown in ball-and-stick models. Leu residues that are important for packing interactions that stabilize the tetramer of LS2 are shown in green sticks. In panel (*b*), the carbon atoms of the porphyrin cofactor are shown in purple. (*c*) Rocker, a *de novo* designed zinc transporter, showing configurations that were used for positive (+) and negative (–) design.

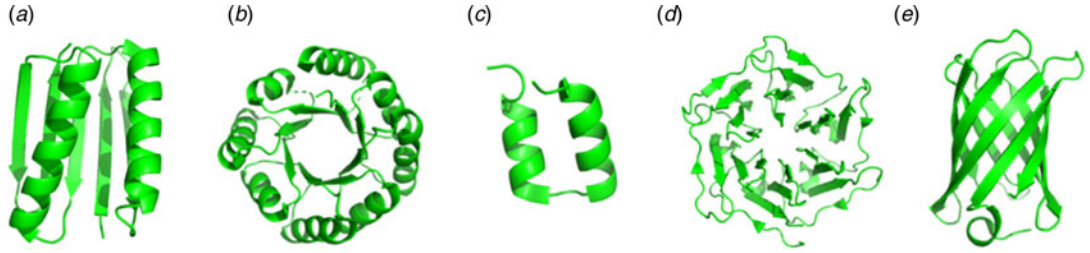


Fig. 11. Representative examples of *de novo* designed protein scaffolds. (*a*) TOP7, a *de novo* designed fold with no natural analogs (PDB: 1QYS). (*b*) A computationally designed TIM barrel (PDB: 5BVL). (*c*) A *de novo* designed mini protein (PDB: 5TX8). (*d*). Pizza6, a *de novo* designed fold with no natural analogs (PDB: 6F0Q). (*e*) A *de novo* designed β-barrel (PDB: 6D0T).

protein that was both rich in β structure and also had a fold not previously seen in nature.

The successful design of TOP7 (Kuhlman *et al.*, 2003) introduced an exciting new chapter, in which backbone fragment libraries derived from the PDB were used to build-up backbones of *de novo* proteins. This approach provided a solution to what was largely a chicken-egg problem in protein design. In *de novo* design, one needs a backbone structure to design a sequence, but it is hard to specify the precise backbone structure without first specifying the sequence. Thus, the designer is set with the task of designing a 'designable' backbone structure (i.e. one that can be stabilized by a sequence composed of the 20 commonly occurring amino acids in this case). Today, one typically tests on the order of 10^3 to 10^5 backbones to see which are designable. For each possible backbone, one uses sidechain repacking and other sequence-design algorithms to determine whether it can be outfitted with a sequence that satisfies the physical restraints required for folding. In rotamer-based approaches to protein sequence selection, a Monte-Carlo algorithm is used to discover sequences that are predicted to fold into the desired structure using a pairwise decomposable potential function that allows an efficient search through amino acid sequence and rotamer space for any given backbone.

The question then is how one specifies a foldable backbone. While significant success was obtained using helical bundles that were specified using a set of algebraic equations, many protein folds are too asymmetric to describe using reduced-parameter models. Kuhlman, Baker, and coworkers introduced an approach that circumvented this problem (Kuhlman *et al.*, 2003). In their approach, one first defines a coarse-grained graph of the desired protein, which contains information such as the positions of secondary structure and inter-residue contacts. This blueprint defines the target fold and guides the search for a foldable backbone that satisfies the design restraints. Mainchain fragments

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

Box 2. Structural bioinformatics, sequence propensities, and fragment-based strategies for *de novo* protein design

Although the first protein design algorithms were based on molecular mechanics force fields, over the years, the scoring and search algorithms have evolved to greater complexity. Modern scoring functions, such as that used in Rosetta now include contributions from physical molecular mechanics force fields, terms to approximate the hydrophobic effect, and residue-specific and sequence-specific mainchain statistics. Additionally, Rosetta uses fragments of up to 15 residues in length to build protein structures, as well as the underlying sequence probabilities to score them energetically. While absent from early approaches to *de novo* protein design, statistical terms and fragment libraries have become increasingly important to enable design of ever-more complex structures. Thus, in modern approaches to protein design, fragments from the PDB are clustered based on 3D similarity and used in assembly procedures for protein structure prediction and design (Leaver-Fay *et al*., 2011; Marcos *et al*., 2017).

The underlying principles that are encoded in structural informatics can be understood and reconciled to physical principles. For example, it has long been known that different amino acid residues have distinct propensities for adopting a given secondary structure of being found in a given environment (e.g. buried *versus* exposed), and the underlying energetics can be roughly approximated through the Boltzmann distribution (Chou and Fasman, 1978). The derived pseudo-energies are generally in good agreement with more direct experiments (Miller *et al*., 1987). Rotameric preferences are in agreement with torsional potentials from molecular mechanics (Dunbrack and Karplus, 1993; Dunbrack and Karplus, 1994) and hence can be extended to design of non-natural foldamers (Shandler *et al*., 2010). Also, the sequence-specific positional preferences for forming β turns or capping helices can be reconciled to first-principles (Wilmot and Thornton, 1988; Efimov, 1993). Similarly, the more residue-specific sequence preferences used in modern design algorithms likely reflect the sequence/energy landscape for a given substructure.

from crystallographic structures are then combined and spliced together to create physically reasonable backbones that also conform to the guiding restraints. Sequences are then designed based on this initial draft of a backbone. In the next step structures are predicted for the designed sequences – again using backbone fragment assembly together with conformational energy minimization to facilitate the backbone search. The design then proceeds through repeated cycles of structure prediction for a given designed sequence and sequence redesign of the resulting predicted structures. Thus, through repeated cycles of sequence design and structure prediction the computation converges on a highly designable structure-backbone combination. Using this approach Kuhlman and Baker designed TOP7, which was highly stable and showed all the characteristics of a native-folded protein. Most importantly, its crystallographic structure was in outstanding agreement (1.2 Å backbone root mean square deviation) with the design. A major milestone in *de novo* protein design had been crossed, with significant implications for the design of proteins with a variety of folds (Box 2).

The extension of the success of TOP7 to other folds required a general platform for inputting blueprints for design and implementing them as restraints in flexible backbone design. This was realized in RosettaRemodel (Huang *et al*., 2011), which provided the framework for a wide range of design problems including: the insertion, deletion, and remodeling of loops; design of disulfides; input of symmetry operators, and various other aspects of *de novo* design. Importantly, parametric approaches to backbone design could be flexibly incorporated into the blueprint to facilitate the entire design process. This framework led to many successes in *de novo* design over the last half decade. Given that this work has been recently reviewed (Huang *et al*., 2016*a*) we will discuss it only briefly here, and focus on a few outstanding examples of Rosetta designs that have appeared in the last 2 years. The design of a number of α-β folds, including rubredoxin, P-loop, and Rossman folds have been achieved (Lin *et al*., 2015; Huang *et al*., 2016*a*; Marcos *et al*., 2017). Moreover, repeat proteins (discussed below), mini-proteins, and cyclic peptides including ones with D-amino acids or unnatural crosslinks have been prepared and structurally verified (Bhardwaj *et al*., 2016; Chevalier *et al*., 2017; Dang *et al*., 2017; Marcos *et al*., 2018) (Fig. 11).

A number of long-standing problems in *de novo* protein design, including the design of stable all-β proteins have been solved in recent years. In ground-breaking work, Huang *et al*. (2016*b*) solved a classical problem (Goraj *et al*., 1990; Tanaka *et al*., 1994; Houbrechts *et al*., 1995; Figueroa *et al*., 2013) of designing a TIM α-β barrel fold (Fig. 11*b*). The pseudo-symmetry of this fold was idealized, resulting in a protein with approximate four-fold symmetry. Next, in 2018 Marcos *et al*. described the principles for controlling the curvature of β-sheets, and applied them to the design of a series of proteins with curved β-sheets topped with α-helices (Marcos *et al*., 2017). Finally, in 2018 the design of a structurally well-defined all-β barrel protein was reported (Lu *et al*., 2018). The successful design focused on a roughly four-fold symmetrical eight-stranded β-barrel, with the overall shape of a hyperboloid. An initial set of 41 designs were constructed using a set of parametric equations, but all were unsuccessful. A careful analysis of the failed designs showed the accumulation of strain along residues that interact across the strands. More nuanced, symmetry-breaking geometries lead to the successful design, whose structure was in good agreement with the design. Next, a site was introduced to bind a fluorophore in a flat planar geometry. This was not accomplished in a single step, but rather after several cycles of computation, experimental evaluation, generation of combinatorial libraries, and experimental screening. Thus, while it was not possible to design the ligand-binding site by computation alone, it was possible to reach this objective through iterative cycles of experiment and computation.

Combining computational design with experimental library screening to achieve function

In another pioneering contribution, Baker and coworkers (Chevalier *et al*., 2017) integrated large-scale computational design, parallel oligonucleotide synthesis, yeast display screening, and next-generation sequencing to create libraries of approximately 40-residue mini-proteins that bind influenza hemagglutinin (HA), a protein located on a surface of the influenza virus,

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

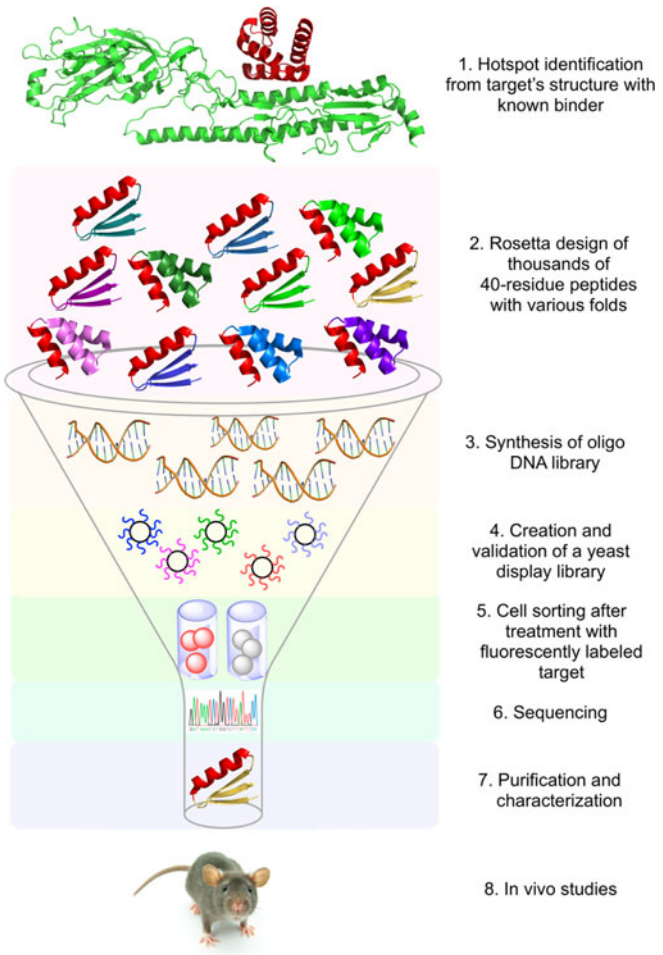


Fig. 12. Overview of the computational design and high-throughput screening of mini-protein binders. Reproduced with permission from Makhlynets and Korendovych (2017). Copyright (2017) American Chemical Society.

and botulinum neurotoxin (BoNT), the most acutely poisonous toxin known. They screened a library of ~4000 backbone geometries representing five well-defined miniprotein folds. 'Hotspot' residues, identified from previous crystal structures of HA and BoNT complexes with different binders, were then grafted onto the mini-protein scaffolds and the rest of sequence was computationally optimized to improve binding and stability. The resulting mini-protein binders were scored based on predicted binding energies and genes encoding ~10 500 of them were synthesized for yeast display. From this pool fluorescence-activated cell sorting enrichment identified 57 and 29 distinct mini-proteins that bound BoNT and HA, respectively. Interestingly, several peptides from a pool of ~6000 scrambled control sequences also showed high affinity for the targets, highlighting the power of naïve libraries in producing strong binders using *de novo* designed scaffolds (Cherny *et al.*, 2012). Importantly, the frequency of finding binders was significantly lower in the randomized sequences. Subsequent refinement of the design model and affinity maturation produced three HA binders and two BoNT binders that displayed excellent affinity (<10 and <1 nM, respectively) for their targets. These K_ds are on par with those of scFv's (~200 pM) derived for the same target and just three orders of magnitude less potent than the corresponding monoclonal antibodies (~320 fM) (Kalb *et al.*, 2010). It is noteworthy that, given the small size of the mini-proteins, in practical terms the same dose by mass results in a comparable therapeutic effect. HB1.6928.2.3, one of the mini-protein binders of HA, provided full protection of mice from influenza before exposure to a lethal dose of the virus and 100% survival after intranasal administration of a single-therapeutic dose 24 h after the exposure.

The multistep approach developed by Baker and coworkers (Fig. 12) that employs protein design to combine the recognition power of 'antibody-like' protein–protein interactions with the stability and ease of production of small, potentially non-immunogenic, mini-proteins allows high-throughput screening to reach its full potential. The demonstrated success rate of finding a strong binder in a computationally designed library (~1%) is sufficient for identification of hits using high-throughput screening.

Design of protein assemblies

Nature has evolved proteins that form structurally complex and functionally rich supramolecular assemblies. Given the wide array of structures that proteins can adopt, targeted design of protein-based supramolecular assemblies can provide a path to novel functional materials and nanostructures. *De novo* design of complex assemblies is rapidly expanding in manifold directions, and several recent reviews of the area are available (De Santis and Ryadnov, 2015; Norn and Andre, 2016; Kobayashi and Arai, 2017; Yeates, 2017; Beesley and Woolfson, 2019). Furthermore, the pallet for design has expanded to include a variety of materials including aromatic peptides and collagen triple helical peptides, which are beyond the scope of the current review. Here, we focus primarily on structures in which designed *de novo* are realized by expansion of the fragment assembly and parametric approaches discussed above.

Elongation in one dimension: superhelical assemblies with translational and screw symmetries

Advances in protein design have enabled the design of complex linear arrays of proteins with diverse morphologies, including fibers and nanotubes that can be used for molecular nanocompartment encapsulation, drug delivery, tissue engineering, and catalysis both *in vivo* and *in vitro*. Work in this field has exploded in recent years, and has been extensively reviewed elsewhere (Zhang *et al.*, 1999; Gazit, 2007; Childers *et al.*, 2009; Aida *et al.*, 2012; Webber *et al.*, 2016; Kumar *et al.*, 2018; Nguyen and Ueno, 2018). Here, we attempt to highlight a few classical, enabling studies as well as recent work, focusing on the principles of design and assembly.

One of the simplest high-order motifs in proteins is the cross-β structure, in which β-strands run perpendicular to the fibril axis, forming infinite parallel or antiparallel sheets (Pauling and Corey, 1951). The sheets often pair along an apolar interface, which further stabilizes the structure (Eisenberg and Jucker, 2012) (Fig. 12). As early as the 1980s (DeGrado and Lear, 1985; Osterman and Kaiser, 1985), simple hydrophobic/hydrophilic patterning was used to design heptapeptides (e.g. LKLKLKL, as discussed above) that can assemble on apolar interfaces into well-defined β-sheet assemblies. These minimalist principles were later extended to design of graphene-binding protofibrils that orient with their main axes along one of six directions defined by graphene's six-fold symmetry (Mustata *et al.*, 2016). Moreover, Zn^{2+}-dependent catalysts with efficiencies that rival those of natural enzymes by weight have also been designed starting

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

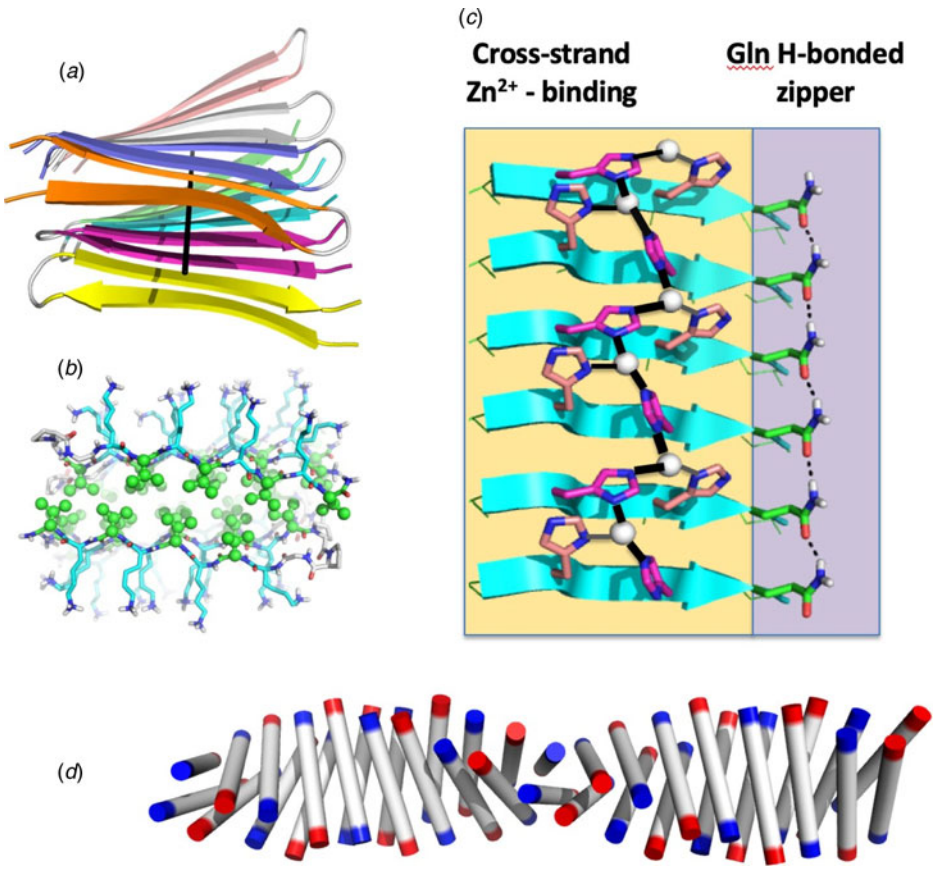


Fig. 13. Structures of amyloid fibrils. (*a*) Strands align perpendicular to the main fibril axis (indicated by a black line) in a structure of MAX1, a strand-turn-strand peptide designed by Schneider and coworkers (PDB: 2N1E). (*b*) Structure of MAX1, with polar Lys residues (blue sticks) on the solvent-exposed surface and apolar Val residues (green ball and sticks) forming a water-free interface. (*c*) Structure of a catalytic Zn^{2+}-binding amyloid (PDB: 5UGK), showing a network of 3-His Zn^{2+} ion coordination, and an H-bonded zipper of Gln sidechains. (*d*) Structure of an α-amyloid assembly, αAm_S (Zhang *et al.*, 2018*b*) (PDB: 6C4Z) the N- and C-termini of the individual helices are designated in blue and red, respectively.

from minimalist principles (Friedmann *et al.*, 2015; Al-Garawi *et al.*, 2017; Zozulia *et al.*, 2018), and the structure of one of these assemblies has been determined by solid state NMR (Fig. 13*c*) (Lee *et al.*, 2017).

Alex Rich and Shuguang Zhang were the first to recognize the potential of amphiphilic β-peptides (Zhang *et al.*, 1993) to form nanofiber scaffolds and membranous structures. Zhang developed such peptides for myriad applications including controlled drug delivery, tissue regeneration, and accelerated wound healing (Zhang, 2017). Similarly, Hamachi has designed a variety of remarkable self-assembling hydrogels that respond to a diverse array of environmental stimuli (Shigemitsu and Hamachi, 2017), and Lynn has used peptide design to explore the possible role of amyloids in early evolution of life (Childers *et al.*, 2009).

Building-up one step in complexity, Schneider designed and determined high-resolution solid-state NMR structures (Nagy-Smith *et al.*, 2015) of fibril-forming peptides consisting of a strand-turn-strand motif (Figs 13*a* and *b*). Members of the MAX1 series of peptides have a range of interesting properties ranging from antimicrobial materials to easily processed hydrogels with finely tuned mechanical properties (Schneider *et al.*, 2002). Shimon, Gazit, and coworkers have built assemblies with similar scaffolds, and characterized their structures by high-resolution X-ray crystallography (Pellach *et al.*, 2017). While the turns in these assemblies connect hydrogen-bonded strands, in other structures such as solenoids and larger amyloids, the turns often connect chains across β-sheets, and these motifs have been successfully used in design of fibrillary assemblies (Pellach *et al.*, 2017).

It is also possible to build fibrous structures from helices rather than β-strands (Fairman and Akerfeldt, 2005). In early work, pioneered by Woolfson, fibrils were built based on staggered pairing interactions between helical coiled-coil peptides; the resulting 'sticky ends' mediated assembly of the peptides into highly elongated coiled coils (Pandya *et al.*, 2000). By introducing kinks or branches they were able to engineer a variety of architectures. Fairman and coworkers adopted a related strategy to induce self-assembly (Fairman and Akerfeldt, 2005; Wagner *et al.*, 2005). These workers designed coiled-coil peptides with insertions that caused the hydrophobic faces to misalign, resulting in a staggered, infinite assembly. More recently, Woolfson have built an orthogonal set of rotationally symmetric coiled coils, ranging from dimers to heptamers, which can be used as building blocks to create a variety of assemblies (Woolfson *et al.*, 2015). By introducing favorable electrostatic, hydrophobic interactions or metal–ligand interactions (Nambiar *et al.*, 2018) near the ends of these coiled coils it is possible to induce assembly into infinite super-helical assemblies with the central axis of the individual coiled coils aligning along the fibril axis (Burgess *et al.*, 2015). When the end-to-end interactions are designed to be highly favorable, the bundles assemble in solution to form elongated fibers (Fig. 14*a*); weaker end-to-end interactions can be used to induce intermonomer contacts in crystals (Ogihara *et al.*, 1997; Lanci *et al.*, 2012; Zhang *et al.*, 2018*b*)

In other examples of helical fibril engineering, the individual helices are designed to form lateral assemblies that run nearly perpendicular to the fibril axis. Conticello and coworkers have designed assemblies of peptides based on a heptad repeat, in which two hydrophobic faces, (*a*/*d*) and (*c*/*f*), are separated by

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

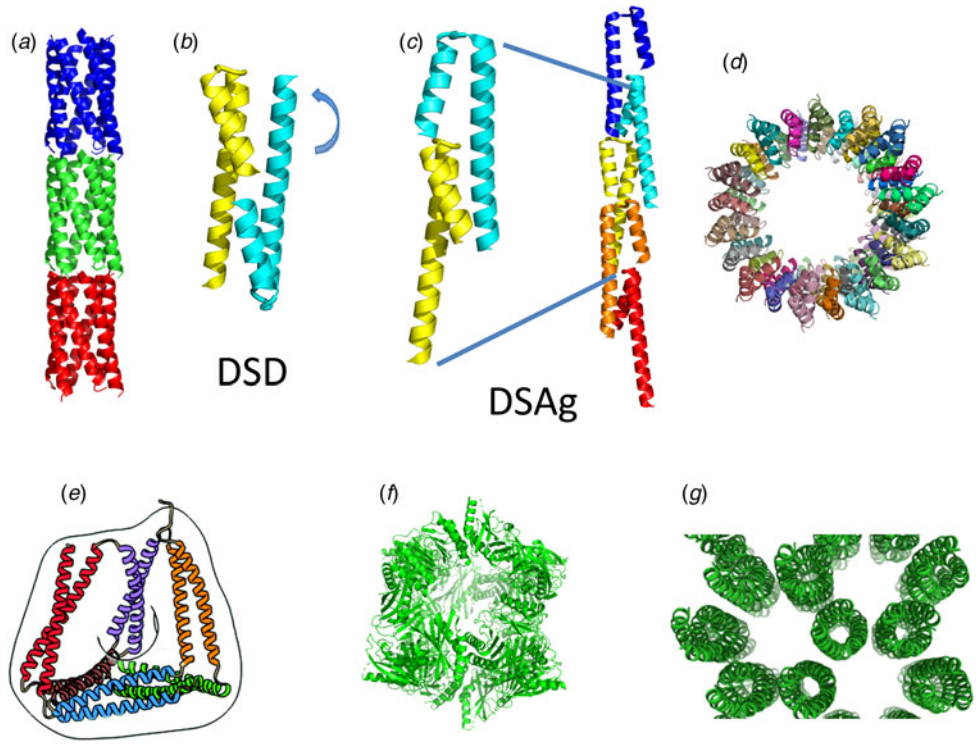


Fig. 14. Structural assemblies of designed proteins. Proteins that assemble in one dimension to form fibers and tubes are shown in panels (*a*–*d*). Panel (*a*) shows the structure of a hexameric bundle designed (PDB: 4H8M), that has been engineered to assemble into stacked bundles (structure inferred by EM). Panel (*b*) illustrates a dimeric three-helix bundle assembled from helix–loop–helix motifs (PDB: 1G6U) consisting of one short and one long helix. The sequence was designed to cause the units to assemble with the loops on opposite sides of the bundle in an 'up-down' orientation to give a domain-swapped dimer. In a second design, the sequence was designed to cause the loops to align in an 'up-up' orientation that induced fibril formation. Panel (*d*) illustrates larger-diameter nano-pores composed of helix–loop–helix motifs (PDB: 6MK1), and panel (*e*) shows the assembly scheme for TET12SN family peptides that spontaneously assemble into a tetrahedral cages (reproduced from Lapenta (2018) 351 – Published by The Royal Society of Chemistry). Panel (*f*) illustrates a tetrahedral protein cage created by computationally designing protein–protein interfaces (PDB: 4NWR), and panel (*g*) illustrates a computationally designed protein crystal (PDB: 4H8M).

polar residues. Cryo-electron microscopy (EM) revealed that the peptides assemble into wide tubes that can encapsulate small molecules (Xu *et al.*, 2013; Shen *et al.*, 2018). Additionally, peptides have been designed to assemble with their axes perpendicular to the fibril axis precisely as in twisted cross-β structures, forming very long-twisted 'cross-α' fibrils as seen in crystallographic structures of the assemblies (Zhang *et al.*, 2018*b*) (Fig. 13*d*). These peptides were used to direct assembly of fused florescent proteins in mammalian cells, and by varying the sequence it was possible to modulate the structure and assembly/disassembly kinetics. Cross-α structures are also of current interest, because they have been discovered in toxic peptides (Tayeb-Fligelman *et al.*, 2017). It will be interesting to see how wide-spread this assembly motif might be.

In each of the above examples of coiled-coil assemblies, the monomeric unit was a single helical peptide. Alternatively, more complex units can be used to create structures with greater structural diversity. The earliest example involved the design of fibrils assembled from domain-swapped versions of a three-helix bundle (Ogihara *et al.*, 2001) related to α3D (Bryson *et al.*, 1995, 1998). The basic design unit consisted of a hairpin consisting of one long and one short helix designed to assemble into three-helix structures (Figs 14*b* and *c*). Electrostatic interactions were manipulated to allow the unit to assemble into a closed, domain-swapped dimer a fibrillar array, depending on whether the helix–loop–helix motifs assembled with the loops in an *anti* (Fig. 14*b*) or a *cis* orientation. X-ray crystallography and EM confirmed the structure of the domain-swapped dimer and fibril, respectively (Ogihara *et al.*, 2001). Furthermore, analysis of the crystallographic structure of a domain-swapped dimer illustrated principles for design of antiparallel six-helix bundles (Ghirlanda *et al.*, 2002). Finally, by redesigning the hydrophobic core of the hexameric bundles, Grigoryan *et al.* engineered bundles that selectively solubilized only a single form of carbon nanotubes (Grigoryan *et al.*, 2011).

In nature, covalently assembled superhelical repeat proteins are often assembled by repeating simple motifs such as helix–loop–helix motifs with intervening tight loops or turns (Kobe and Kajava, 2000; Kajava, 2012). Consensus sequence motifs have been generated for repeat proteins, and used to create robust scaffolds for selection of peptide binding proteins (Kajander *et al.*, 2006; Pluckthun, 2015). A number of repeats are assembled into a single-protein chain, and N- and C-terminal 'capping motifs' are also included to avoid run-away non-covalent assembly into fibrils. Taking a different approach, Conticello and coworkers used non-covalent assembly of peptides patterned after the helix–loop–helix motifs of thermophilic HEAT and leucine-rich variant motifs (Fig. 14*c*). Cryo-EM structures at near-atomic resolution demonstrated the formation of tubes with outer radii of 70 or 80 Å (Hughes *et al.*, 2019).

Recently, André and coworkers took a structure-based approach to design repeat proteins based on the leucine-rich repeat. They built on the known structures of natural proteins to design repeats with predefined shapes, which are then

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

assembled to create helical arrays with predetermined superhelical geometries (Ramisch *et al.*, 2014). ElGamacy, Lupas, and coworkers used an interface-directed strategy to design less regular solenoid-like proteins, in which the helix–loop–helix motifs alternating the handedness rather than repeating with exact symmetry (ElGamacy *et al.*, 2018).

Baker and coworkers have developed generalized computational methods to engineer cyclic and superhelical arrays formed from helix–loop–helix motifs (Brunette *et al.*, 2015; Doyle *et al.*, 2015). The Rosetta program is used to build repeating helix–loop–helix–loop motifs, in which the backbone and sequence is identical for each repeat. This procedure generates well-packed superhelical repeat proteins; the desired superstructure can be specified by adding a pseudo-energy term that penalizes for geometries that do not match the desired superhelical curvature and rise. Using this method, Brunette *et al.* explored the structure space for helical repeat proteins containing a range of helix–loop–helix geometries (Brunette *et al.*, 2015). The resulting structures have been verified at atomic resolution, and a number of geometries not yet seen in crystal structures of natural proteins were designed and experimentally demonstrated. These methods have been extended to the design of filamentous arrays formed from previously characterized *de novo*-designed helical bundles (Shen *et al.*, 2018). Using Rosetta, a variety of well packed motifs are sampled and replicated to create a range of superhelical geometries. Of 124 designs tested, 34 formed filaments – six of which were structurally characterized and found to agree with the underlying design to varying degree of accuracy.

Elongation in two dimensions: planar lattice-like structures

The design of lattice-like structures can be realized by *de novo* designed protein into a unit cell, and arranging its orientation and sequence to create a stable assembly. The first structurally verified *de novo* design of a two-dimensional (2D) assembly focused on P321 and P6 arrays of three-helix bundles (Lanci *et al.*, 2012). This work has been expanded to design arrays based on tetrameric bundles (Zhang *et al.*, 2016). Both examples employed the SCAD sequence design algorithm to generate the sequence. In each case, the predicted models were in outstanding agreement with the experimental structures. More recently, similar methods have been used by Baker *et al.*, to design a number of different lattices, in this case using natural proteins with cyclic symmetry as the basic building blocks (Gonen *et al.*, 2015). Together, these studies demonstrate the ability to design with Ångstrom-level accuracy over length scales on the order of tens to hundreds of nanometers.

Assembly of cages by combining multiple symmetry elements

The predictable nature and robustness of coiled-coil assemblies was expanded to form large cages (Fletcher *et al.*, 2013) as well as distinct supramolecular polyhedral nanostructures that can assemble both *in vitro* and *in vivo* (Gradišar *et al.*, 2013; Ljubetič *et al.*, 2017; Park *et al.*, 2017). Marsh and coworkers developed a flexible, symmetry directed approach for creating protein cages by fusing coiled-coil forming peptides to a natural trimeric protein (Sciore *et al.*, 2016). It is also possible to build polyhedral using natural homo-oligomeric proteins as building blocks. Yates and coworkers (Padilla *et al.*, 2001; Yeates, 2017) described general principles for the design of symmetrical virus-like assemblies that form large molecular cages by rigid fusions of two oligomeric proteins – for example, one that forms a C_2 dimer with one that forms a C_3 trimer – so that the symmetry axes match the symmetry axes of Euclidean solids. This method has been used to create novel cages with varying symmetries including icosahedral assemblies (King *et al.*, 2012; Bale *et al.*, 2016; Hsia *et al.*, 2016). Baker and coworkers have implemented and extended this approach to allow design of large, well-defined, virus-like protein cages with atomic accuracy, including proteins capable of encapsulating their own DNA (Butterfield *et al.*, 2017). Hilvert and coworkers further modified these computationally designed protein cages to deliver oligonucleotides to efficiently regulate gene expression in mammalian cells (Edwardson and Hilvert, 2019).

Elongation in three dimensions: crystal engineering

Progressing from the design of 2D arrays to macroscopic 3D crystals represents the highest level of complexity. Conceptually, this can be achieved by engineering the assembly of a 2D lattice (e.g. as discussed above in the section 'Elongation in two dimensions: planar lattice-like structures') into a third dimension. However, designing predetermined crystal structures is subtle, given the size and complexity of proteins and the myriad noncovalent interactions that govern protein crystallization. Saven, DeGrado, and coworkers developed a computational approach to design a helical bundle that assembles in P6, a polar, layered crystallographic space group with both C_2 and C_3 symmetry axes (Lanci *et al.*, 2012). A C_3-symmetric helical bundle was placed along the three-fold axis, and its orientation and unit cell parameters were systematically varied to create a sequence-structure energy landscape using the SCADS program for computational protein design. A hierarchy of interactions of graded stability was used in the design. Strongly stabilizing hydrophobic and packing interactions were engineered to stabilize the core of the three-helix bundle, while weaker packing interactions between surface-exposed Gly and Ala interactions were used to stabilize lateral interactions between the helices. Finally, end-to-end hydrogen bonds between helical ends stabilized the stacking of columns of helical bundles. A 2.1 Å resolution X-ray crystal structure of one such designed protein exhibits sub-Ångstrom agreement with the computational model in the spacing and parallel ordering of neighboring proteins in the crystal. The crystals have large hexagonal channels, which should be able to accommodate a variety of small- to meso-sized molecular cargos. For example, similar crystals of designed coiled coils have been found to organize C_{60} derivatives into arrays with interesting electronic properties (Kim *et al.*, 2016).

Summary and outlook

In the past several decades, the design of *de novo* proteins with predetermined structures and functions has progressed from an outrageous concept to a routine accomplishment, with far-reaching implications for the fields of chemistry, nanoscience, and biotechnology. *De novo* design is a compellingly critical test of our understanding of protein structure and function. If we understand proteins we should be able to design them from scratch. This approach translates our passive understanding of proteins to an active understanding that is already enabling the design of proteins and biomimetic polymers with properties not available in nature.

The first grand challenge our field encountered was the protein folding problem – how does an amino acid sequence code for the

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

3D structure of a protein? Today, we understand the principles of protein folding sufficiently well to design proteins with a large range of sizes, sizes, dynamic properties and with thermodynamic stabilities far exceeding those seen in nature. Given this ability to control tertiary structure, protein designers are also tackling the problem of designing function. Initial work in this area has been primarily fundamental, as we have progressed from a passive understanding to the active understanding needed to design functional proteins from scratch. Nevertheless, practical applications have already emerged and clearly will expand. In this review, we focused on three functions: binding, catalysis, and vectoral transport through membranes. Sufficient progress has been made in each to reasonably extrapolate what we might reasonably expect to achieve in the next decade.

The first clearly defined achievements in the area of binding focused on selective and geometrically specific recognition of transition metal ions. The initial designs focused on binding of metal ions in relatively stable, common geometries, as in structural metal sites in proteins. With time, *de novo* design proteins were produced with more interesting metal sites capable of catalyzing a variety of oxidative, reductive, and hydrolytic processes. Thus, *de novo* design is now increasingly used to understand how proteins influence the reactivity and catalytic properties of their metal ion cofactors. Furthermore, a large number of proteins have been designed to bind non-biological metal ions and metal-organic complexes in precisely predetermined structures and environments. These accomplishments raise the possibility of designing cofactor-containing proteins for diverse applications ranging from optical devices to catalysts that combine the advantages of traditional transition metal catalysts with the versatility, programmability, and water solubility of proteins.

A second binding functionality that has been achieved involves the design of peptides and proteins that bind to protein interfaces. *De novo* design methods have enabled the design of proteins that are smaller and much more stable toward chemical, enzymatic, and thermal denaturation than natural proteins such as antibodies. *De novo* design is also providing increasingly good starting points for experimental optimization of binding affinity and specificity. *De novo* designed proteins have considerable potential as therapeutics for pharmaceutical intervention of unmet medical needs.

The design of proteins that bind complex, highly functionalized small molecules remains a larger challenge that has only now being addressed . The design of small molecule binders requires mastery of some of the most difficult problems in protein design. First, a binding cavity must be constructed to encompass the molecule of interest. In early studies where this was accomplished (Di Costanzo *et al.*, 2001; Lombardi *et al.*, 2001; Geremia *et al.*, 2005; Lombardi *et al.*, 2019), building a small-molecule binding site was very destabilizing to the protein conformation and required careful optimization of other regions of the tertiary structure (Faiella *et al.*, 2009). Once, a cavity has been constructed, the designed protein must also position polar sidechains appropriately to form highly directional hydrogen-bonded interactions to the ligand (in cases where the binding of densely functionalized polar ligands is desired). Finally, when the target small molecule contains a number of rotatable bonds, the ligand–protein interactions need to be highly favorable to compensate for the unfavorable entropy associated with binding the small molecule in a single conformation. While challenging, we expect that advances in sampling and scoring ligand–protein poses will enable successful design of small molecule-binding proteins without the need for repeated cycles of experimental optimization. The attainment of this ability will be an important step in the design of proteins that catalyze kinetically challenging reactions with efficiencies approaching those of natural enzymes.

We have also seen significant progress in the design of proteins that assemble in membranes and other non-aqueous or heterogeneous environments. It is now possible to design membrane proteins and assemblies with very high stabilities and predictable structures. We have also seen the first examples of proteins that facilitate transport of electrons and polar solutes across phospholipid bilayers. Applications of such systems to single-molecule sensing are likely to follow. For example, highly engineered variants of natural proteins are now used for sequencing RNA and DNA using the nanopore technology (Branton *et al.*, 2008). It will be exciting to construct proteins from scratch for such demanding applications.

Finally, methods for protein design are developing very rapidly. In this review, we saw that the earliest proteins were designed using simple physical principles and molecular mechanics force fields. More recently developed methods increasingly rely on backbone fragments and statistical quantities derived from structural bioinformatics to sample foldable protein structures and sequences. Nevertheless, the same physical principles are involved and incorporated into modern force fields for protein design. In the coming years, there will doubtlessly be improvements in both approaches. Advances in computing will allow all-atom molecular dynamics calculations using both implicit and explicit solvents at various steps within the design workflow. Such methods will allow one to better model non-canonical structures and to evaluate the potential success of designs. In parallel, the power of bioinformatics will increase dramatically with the inclusion of machine learning (Mackenzie *et al.*, 2016; Mackenzie and Grigoryan, 2017; Eguchi and Huang, 2019). Advanced non-supervised approaches will enable one to discover highly favorable atomic arrangements that are difficult to sample with high precision and quantify with current methods. Machine-learning methods will contribute to the identification of stable 'designable' tertiary structures that can be designed using the 20 commonly occurring amino acids. Generative adversarial networks will be used to generate both tertiary structures and sequences starting with only a rough draft of the desired structure.

In summary, *de novo* protein design has evolved into a vibrant approach for testing hypotheses concerning the fundamental aspects of protein folding and function, and it is now brimming with potential for applications in sensing, catalysis, pharmaceuticals, and nanotechnology. Given recent improvements in computing, including advanced methods for machine learning, one can expect advances to accelerate dramatically in the coming years.

Acknowledgements. The authors thank he NIH (Grants GM122603 to W.F.D. and GM119634 to I.V.K.), the NSF (Grant 1709506 to W.F.D.) and the CRDF (Grant No. OISE-18-63891-0 to I.V.K.) for support.

References

Adamian L and Liang J (2001) Helix–helix packing and interfacial pairwise interactions of residues in membrane proteins. *Journal of Molecular Biology* **311**, 891–907.

Adamian L and Liang J (2002) Interhelical hydrogen bonds and spatial motifs in membrane proteins: polar clamps and serine zippers. *Proteins* **47**, 209–218.

Adhikari AN, Freed KF and Sosnick TR (2012) *De novo* prediction of protein folding pathways and structure using the principle of sequential

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

stabilization. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 17442–17447.

Aida T, Meijer EW and Stupp SI (2012) Functional supramolecular polymers. *Science* **335**, 813–817.

Åkerfeldt K, Kim RM, Camac D, Groves JT, Lear JD and DeGrado WF (1992) Tetraphilin: a four-helix proton channel built on a tetraphenylporphyrin framework. *Journal of the American Chemical Society* **114**, 9656–9657.

Åkerfeldt KS, Lear JD, Waserman ZR, Chung LA and DeGrado WF (1993) Synthetic peptides as models for ion channel proteins. *Accounts of Chemical Research* **26**, 191–197.

Al-Garawi ZS, McIntosh BA, Neill-Hall D, Hatimy AA, Sweet SM, Bagley MC and Serpell LC (2017) The amyloid architecture provides a scaffold for enzyme-like catalysts. *Nanoscale* **9**, 10773–10783.

Allert M and Baltzer L (2003) Noncovalent binding of a reaction intermediate by a designed helix–loop–helix motif-implications for catalyst design. *ChemBioChem* **4**, 306–318.

Anderson SM, Mueller BK, Lange EJ and Senes A (2017) Combination of Calpha-H hydrogen bonds and van der Waals packing modulates the stability of GxxxG-mediated dimers in membranes. *Journal of the American Chemical Society* **139**, 15774–15783.

Arbely E and Arkin IT (2004) Experimental measurement of the strength of a C alpha-H…O bond in a lipid bilayer. *Journal of the American Chemical Society* **126**, 5362–5363.

Argos P, Rossmann MG and Johnson JE (1977) A four-helical supersecondary structure. *Biochemical and Biophysical Research Communications* **75**, 83–86.

Baker RP and Urban S (2012) Architectural and thermodynamic principles underlying intramembrane protease function. *Nature Chemical Biology* **8**, 759–768.

Bale JB, Gonen S, Liu Y, Sheffler W, Ellis D, Thomas C, Cascio D, Yeates TO, Gonen T, King NP and Baker D (2016) Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389–394.

Baltzer L (2011) Crossing borders to bind proteins – a new concept in protein recognition based on the conjugation of small organic molecules or short peptides to polypeptides from a designed set. *Analytical and Bioanalytical Chemistry* **400**, 1653–1664.

Baltzer L, Broo KS, Nilsson H and Nilsson J (1999) Designed four-helix bundle catalysts – the engineering of reactive sites for hydrolysis and transesterification reactions of p-nitrophenyl esters. *Bioorganic & Medicinal Chemistry* **7**, 83–91.

Beesley JL and Woolfson DN (2019) The *de novo* design of alpha-helical peptides for supramolecular self-assembly. *Current Opinion in Biotechnology* **58**, 175–182.

Bender GM, Lehmann A, Zou H, Cheng H, Fry HC, Engel D, Therien MJ, Blasie JK, Roder H, Saven JG and DeGrado WF (2007) *De novo* design of a single-chain di phenyl porphyrin metalloprotein. *Journal of the American Chemical Society* **129**, 10732–10740.

Berwick MR, Lewis DJ, Jones AW, Parslow RA, Dafforn TR, Cooper HJ, Wilkie J, Pikramenou Z, Britton MM and Peacock AF (2014) *De novo* design of Ln(III) coiled coils for imaging applications. *Journal of the American Chemical Society* **136**, 1166–1169.

Berwick MR, Slope LN, Smith CF, King SM, Newton SL, Gillis RB, Adams GG, Rowe AJ, Harding SE, Britton MM and Peacock AFA (2016) Location dependent coordination chemistry and MRI relaxivity, in *de novo* designed lanthanide coiled coils. *Chemical Science* **7**, 2207–2216.

Betz SF and DeGrado WF (1996) Controlling topology and native-like behavior of *de novo*-designed peptides: design and characterization of antiparallel four-stranded coiled coils. *Biochemistry* **35**, 6955–6962.

Betz SF, Bryson JW, Passador MC, Brown RJ, O'Neil KT and DeGrado WF (1996) Expression of *de novo* designed alpha-helical bundles. *Acta Chemica Scandinavica* **50**, 688–696.

Bhardwaj G, Mulligan VK, Bahl CD, Gilmore JM, Harvey PJ, Cheneval O, Buchko GW, Pulavarti SV, Kaas Q, Eletsky A, Huang PS, Johnsen WA, Greisen PJ, Rocklin GJ, Song Y, Linsky TW, Watkins A, Rettie SA, Xu X, Carter LP, Bonneau R, Olson JM, Coutsias E, Correnti CE, Szyperski T, Craik DJ and Baker D (2016) Accurate *de novo* design of hyperstable constrained peptides. *Nature* **538**, 329–335.

Boyken SE, Chen Z, Groves B, Langan RA, Oberdorfer G, Ford A, Gilmore JM, Xu C, DiMaio F, Pereira JH, Sankaran B, Seelig G, Zwart PH and Baker D (2016) *De novo* design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* **352**, 680–687.

Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, Di Ventra M, Garaj S, Hibbs A, Huang X, Jovanovich SB, Krstic PS, Lindsay S, Ling XS, Mastrangelo CH, Meller A, Oliver JS, Pershin YV, Ramsey JM, Riehn R, Soni GV, Tabard-Cossa V, Wanunu M, Wiggin M and Schloss JA (2008) The potential and challenges of nanopore sequencing. *Nature Biotechnology* **26**, 1146–1153.

Broo KS, Brive L, Ahlberg P and Baltzer L (1997) Catalysis of hydrolysis and transesterification reactions of p-nitrophenyl esters by a designed helix–loop–helix dimer. *Journal of the American Chemical Society* **119**, 11362–11372.

Broo KS, Nilsson H, Nilsson J, Flodberg A and Baltzer L (1998) Cooperative nucleophilic and general-acid catalysis by the HisH(+)-His pair and arginine transition state binding in catalysis of ester hydrolysis reactions by designed helix–loop–helix motifs. *Journal of the American Chemical Society* **120**, 4063–4068.

Brosig B and Langosch D (1998) The dimerization motif of the glycophorin A transmembrane segment in membranes: importance of glycine residues. *Protein Science* **7**, 1052–1056.

Brunette TJ, Parmeggiani F, Huang PS, Bhabha G, Ekiert DC, Tsutakawa SE, Hura GL, Tainer JA and Baker D (2015) Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584.

Bryngelson JD, Onuchic JN, Socci ND and Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167–195.

Bryson JW, Betz SF, Lu HS, Suich DJ, Zhou HX, O'Neil KT and DeGrado WF (1995) Protein design: a hierarchic approach. *Science* **270**, 935–941.

Bryson JW, Desjarlais JR, Handel TM and DeGrado WF (1998) From coiled coils to small globular proteins: design of a native-like three-helix bundle. *Protein Science* **7**, 1404–1414.

Burgess NC, Sharp TH, Thomas F, Wood CW, Thomson AR, Zaccai NR, Brady RL, Serpell LC and Woolfson DN (2015) Modular design of self-assembling peptide-based nanotubes. *Journal of the American Chemical Society* **137**, 10554–10562.

Burton AJ, Thomson AR, Dawson WM, Brady RL and Woolfson DN (2016) Installing hydrolytic activity into a completely *de novo* protein framework. *Nature Chemistry* **8**, 837–844.

Butterfield GL, Lajoie MJ, Gustafson HH, Sellers DL, Nattermann U, Ellis D, Bale JB, Ke S, Lenz GH, Yehdego A, Ravichandran R, Pun SH, King NP and Baker D (2017) Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature* **552**, 415–420.

Calhoun JR, Kono H, Lahr S, Wang W, DeGrado WF and Saven JG (2003) Computational design and characterization of a monomeric helical dinuclear metalloprotein. *Journal of Molecular Biology* **334**, 1101–1115.

Caputo GA and London E (2003) Cumulative effects of amino acid substitutions and hydrophobic mismatch upon the transmembrane stability and conformation of hydrophobic alpha-helices. *Biochemistry* **42**, 3275–3285.

Caputo GA, Litvinov RI, Li W, Bennett JS, DeGrado WF and Yin H (2008) Computationally designed peptide inhibitors of protein–protein interactions in membranes. *Biochemistry* **47**, 8600–8606.

Chakraborty S, Touw DS, Peacock AF, Stuckey J and Pecoraro VL (2010) Structural comparisons of apo- and metalated three-stranded coiled coils clarify metal binding determinants in thiolate containing designed peptides. *Journal of the American Chemical Society* **132**, 13240–13250.

Chakraborty S, Kravitz JY, Thulstrup PW, Hemmingsen L, DeGrado WF and Pecoraro VL (2011) Design of a three-helix bundle capable of binding heavy metals in a triscysteine environment. *Angewandte Chemie (International Edition)* **50**, 2049–2053.

Chen Z, Boyken SE, Jia M, Busch F, Flores-Solis D, Bick MJ, Lu P, VanAernum ZL, Sahasrabuddhe A, Langan RA, Bermeo S, Brunette TJ, Mulligan VK, Carter LP, DiMaio F, Sgourakis NG, Wysocki VH and Baker D (2019) Programmable design of orthogonal protein heterodimers. *Nature* **565**, 106–111.

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

Cherny I, Korolev M, Koehler AN and Hecht MH (2012) Proteins from an unevolved library of *de novo* designed sequences bind a range of small molecules. *ACS Synthetic Biology* **1**, 130–138.

Chevalier A, Silva DA, Rocklin GJ, Hicks DR, Vergara R, Murapa P, Bernard SM, Zhang L, Lam KH, Yao G, Bahl CD, Miyashita SI, Goreshnik I, Fuller JT, Koday MT, Jenkins CM, Colvin T, Carter L, Bohn A, Bryan CM, Fernandez-Velasco DA, Stewart L, Dong M, Huang X, Jin R, Wilson IA, Fuller DH and Baker D (2017) Massively parallel *de novo* protein design for targeted therapeutics. *Nature* **550**, 74–79.

Childers WS, Ni R, Mehta AK and Lynn DG (2009) Peptide membranes in chemical evolution. *Current Opinion in Chemical Biology* **13**, 652–659.

Chino M, Zhang SQ, Pirro F, Leone L, Maglio O, Lombardi A and DeGrado WF (2018) Spectroscopic and metal binding properties of a *de novo* metalloprotein binding a tetrazinc cluster. *Biopolymers* **109**, e23339.

Chiti F and Dobson CM (2009) Amyloid formation by globular proteins under native conditions. *Nature Chemical Biology* **5**, 15–22.

Choma CT, Lear JD, Nelson MJ, Dutton PL, Robertson DE and DeGrado WF (1994) Design of a heme-binding four-helix bundle. *Journal of the American Chemical Society* **116**, 856–865.

Choma C, Gratkowski H, Lear JD and DeGrado WF (2000) Asparagine-mediated self-association of a model transmembrane helix. *Nature Structural & Molecular Biology* **7**, 161–166.

Chou PY and Fasman GD (1978) Empirical predictions of protein conformation. *Annual Review of Biochemistry* **47**, 251–276.

Chung HS, Piana-Agostinetti S, Shaw DE and Eaton WA (2015) Structural origin of slow diffusion in protein folding. *Science* **349**, 1504–1510.

Cochran AG, Skelton NJ and Starovasnik MA (2001) Tryptophan zippers: stable, monomeric beta-hairpins. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5578–5583.

Crichton RR (2019) Biological inorganic chemistry: a new introduction to molecular structure and function. In Crichton R (ed). *Biological Inorganic Chemistry*, 3rd Edn. London, UK: Academic Press.

Crick FHC (1953) The Fourier transform of a coiled-coil. *Acta Crystallographica* **6**, 685–689.

Dahiyat BI and Mayo SL (1996) Protein design automation. *Protein Science* **5**, 895–903.

Dahiyat BI and Mayo SL (1997) *De novo* protein design: fully automated sequence selection. *Science* **278**, 82–87.

Dang B, Wu H, Mulligan VK, Mravic M, Wu Y, Lemmin T, Ford A, Silva DA, Baker D and DeGrado WF (2017) *De novo* design of covalently constrained mesosize protein scaffolds with unique tertiary structures. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 10852–10857.

Das C, Raghothama S and Balaram P (1998) A designed three-stranded beta sheet peptide as a multiple beta-hairpin model. *Journal of the American Chemical Society* **120**, 5812–5813.

DeGrado WF and Lear JD (1985) Induction of peptide conformation at apolar/water interfaces: a study with model peptides of defined hydrophobic periodicity. *Journal of the American Chemical Society* **107**, 7684.

DeGrado WF, Regan L and Ho SP (1987) The design of a four-helix bundle protein. *Cold Spring Harbor Symposia on Quantitative Biology* **52**, 521–526.

DeGrado WF, Wasserman ZR and Lear JD (1989) Protein design, a minimalist approach. *Science* **243**, 622–628.

DeGrado WF, Summa CM, Pavone V, Nastri F and Lombardi A (1999) *De novo* design and structural characterization of proteins and metalloproteins. *Annual Review of Biochemistry* **68**, 779–819.

DeGrado WF, Di Costanzo L, Geremia S, Lombardi A, Pavone V and Randaccio L (2003) Sliding helix and change of coordination geometry in a model di-MnII protein. *Angewandte Chemie (International Edition)* **42**, 417–420.

Der BS, Edwards DR and Kuhlman B (2012*a*) Catalysis by a *de novo* zinc-mediated protein interface: implications for natural enzyme evolution and rational enzyme engineering. *Biochemistry* **51**, 3933–3940.

Der BS, Machius M, Miley MJ, Mills JL, Szyperski T and Kuhlman B (2012*b*) Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. *Journal of the American Chemical Society* **134**, 375–385.

De Santis E and Ryadnov MG (2015) Peptide self-assembly for nanomaterials: the old new kid on the block. *Chemical Society Reviews* **44**, 8288–8300.

Desjarlais JR and Handel TM (1995) *De novo* design of the hydrophobic cores of proteins. *Protein Science* **4**, 2006–2018.

Desmet J, De Maeyer M, Hazes B and Lasters I (1992) The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* **356**, 539–542.

Di Costanzo L, Wade H, Geremia S, Randaccio L, Pavone V, DeGrado WF and Lombardi A (2001) Toward the *de novo* design of a catalytically active helix bundle: a substrate-accessible carboxylate-bridged dinuclear metal center. *Journal of the American Chemical Society* **123**, 12749–12757.

Dieckmann GR, McRorie DK, Tierney DL, Utschig LM, Singer CP, O'Halloran TV, Penner-Hahn JE, DeGrado WF and Pecoraro VL (1997) *De novo* design of mercury-binding two- and three-helical bundles. *Journal of the American Chemical Society* **119**, 6195–6196.

Dieckmann GR, McRorie DK, Lear JD, Sharp KA, DeGrado WF and Pecoraro VL (1998) The role of protonation and metal chelation preferences in defining the properties of mercury-binding coiled coils. *Journal of Molecular Biology* **280**, 897–912.

Dieckmann GR, Lear JD, Zhong Q, Klein ML, DeGrado WF and Sharp KA (1999) Exploration of the structural features defining the conduction properties of a synthetic ion channel. *Biophysical Journal* **76**, 618–630.

Dou J, Vorobieva AA, Sheffler W, Doyle LA, Park H, Bick MJ, Mao B, Foight GW, Lee MY, Gagnon LA, Carter L, Sankaran B, Ovchinnikov S, Marcos E, Huang PS, Vaughan JC, Stoddard BL and Baker D (2018) *De novo* design of a fluorescence-activating beta-barrel. *Nature* **561**, 485–491.

Doura AK, Kobus FJ, Dubrovsky L, Hibbard E and Fleming KG (2004) Sequence context modulates the stability of a GxxxG-mediated transmembrane helix–helix dimer. *Journal of Molecular Biology* **341**, 991–998.

Doyle L, Hallinan J, Bolduc J, Parmeggiani F, Baker D, Stoddard BL and Bradley P (2015) Rational design of alpha-helical tandem repeat proteins with closed architectures. *Nature* **528**, 585–588.

Dunbrack Jr RL and Cohen FE (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Science* **6**, 1661–1681.

Dunbrack Jr RL and Karplus M (1993) Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *Journal of Molecular Biology* **230**, 543–574.

Dunbrack Jr RL and Karplus M (1994) Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Natural Structural Biology* **1**, 334–340.

Duong MT, Jaszewski TM, Fleming KG and MacKenzie KR (2007) Changes in apparent free energy of helix–helix dimerization in a biological membrane due to point mutations. *Journal of Molecular Biology* **371**, 422–434.

Duran AM and Meiler J (2018) Computational design of membrane proteins using RosettaMembrane. *Protein Science* **27**, 341–355.

Eck RV and Dayhoff MO (1966) Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences. *Science* **152**, 363–366.

Edwardson TGW and Hilvert D (2019) Virus-inspired function in engineered protein cages. *Journal of the American Chemical Society* **141**, 9432–9443.

Efimov AV (1993) Patterns of loop regions in proteins. *Current Opinion in Structural Biology* **3**, 379–384.

Eguchi RR and Huang PS (2019) Multi-scale structural analysis of proteins by deep semantic segmentation. *Bioinformatics (Oxford, England)*, btz560.

Eilers M, Shekar SC, Shieh T, Smith SO and Fleming PJ (2000) Internal packing of helical membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5796–5801.

Eisenberg D and Jucker M (2012) The amyloid state of proteins in human diseases. *Cell* **148**, 1188–1203.

Eisenberg D, Wilcox W, Eshita SM, Pryciak PM, Ho SP and DeGrado WF (1986) The design, synthesis, and crystallization of an alpha-helical peptide. *Proteins* **1**, 16–22.

Elazar A, Weinstein J, Biran I, Fridman Y, Bibi E and Fleishman SJ (2016) Mutational scanning reveals the determinants of protein insertion and association energetics in the plasma membrane. *Elife* **5**, e12125

ElGamacy M, Coles M, Ernst P, Zhu H, Hartmann MD, Pluckthun A and Lupas AN (2018) An interface-driven design strategy yields a novel, corrugated protein architecture. *ACS Synthetic Biology* **7**, 2226–2235.

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

Emberly EG, Mukhopadhyay R, Tang C and Wingreen NS (2004) Flexibility of beta-sheets: principal component analysis of database protein structures. *Proteins* **55**, 91–98.

Engelman DM, Chen Y, Chin CN, Curran AR, Dixon AM, Dupuy AD, Lee AS, Lehnert U, Matthews EE, Reshetnyak YK, Senes A and Popot JL (2003) Membrane protein folding: beyond the two stage model. *FEBS Letters* **555**, 122–125.

Faiella M, Andreozzi C, de Rosales RT, Pavone V, Maglio O, Nastri F, DeGrado WF and Lombardi A (2009) An artificial di-iron oxo-protein with phenol oxidase activity. *Nature Chemical Biology* **5**, 882–884.

Fairman R and Akerfeldt KS (2005) Peptides as novel smart materials. *Current Opinion in Structural Biology* **15**, 453–463.

Fersht AR and Serrano L (1993) Principles of protein stability derived from protein engineering experiments. *Current Opinion in Structural Biology* **3**, 75–83.

Figueroa M, Oliveira N, Lejeune A, Kaufmann KW, Dorr BM, Matagne A, Martial JA, Meiler J and Van de Weerdt C (2013) Octarellin VI: using Rosetta to design a putative artificial (beta/alpha)8 protein. *PLoS One* **8**, e71858.

Fletcher JM, Harniman RL, Barnes FR, Boyle AL, Collins A, Mantell J, Sharp TH, Antognozzi M, Booth PJ, Linden N, Miles MJ, Sessions RB, Verkade P and Woolfson DN (2013) Self-assembling cages from coiled-coil peptide modules. *Science* **340**, 595–599.

Friedmann MP, Torbeev V, Zelenay V, Sobol A, Greenwald J and Riek R (2015) Towards prebiotic catalytic amyloids using high throughput screening. *PLoS One* **10**, e0143948.

Fry HC, Lehmann A, Saven JG, DeGrado WF and Therien MJ (2010) Computational design and elaboration of a *de novo* heterotetrameric alpha-helical protein that selectively binds an emissive abiological (porphinato) zinc chromophore. *Journal of the American Chemical Society* **132**, 3997–4005.

Fry HC, Lehmann A, Sinks LE, Asselberghs I, Tronin A, Krishnan V, Blasie JK, Clays K, DeGrado WF, Saven JG and Therien MJ (2013) Computational *de novo* design and characterization of a protein that selectively binds a highly hyperpolarizable abiological chromophore. *Journal of the American Chemical Society* **135**, 13914–13926.

Fujiwara D and Fujii I (2013) Phage selection of peptide 'microantibodies'. *Current Protocols in Chemical Biology* **5**, 171–194.

Fujiwara D, Kitada H, Oguri M, Nishihara T, Michigami M, Shiraishi K, Yuba E, Nakase I, Im H, Cho S, Joung JY, Kodama S, Kono K, Ham S and Fujii I (2016) A cyclized helix–loop–helix peptide as a molecular scaffold for the design of inhibitors of intracellular protein–protein interactions by epitope and arginine grafting. *Angewandte Chemie (International Edition)* **55**, 10612–10615.

Gadzala M, Dulak D, Kalinowska B, Baster Z, Brylinski M, Konieczny L, Banach M and Roterman I (2019) The aqueous environment as an active participant in the protein folding process. *Journal of Molecular Graphics & Modelling* **87**, 227–239.

Gazit E (2007) Self-assembled peptide nanostructures: the design of molecular building blocks and their technological utilization. *Chemical Society Reviews* **36**, 1263–1269.

Geremia S, Di Costanzo L, Randaccio L, Engel DE, Lombardi A, Nastri F and DeGrado WF (2005) Response of a designed metalloprotein to changes in metal ion coordination, exogenous ligands, and active site volume determined by X-ray crystallography. *Journal of the American Chemical Society* **127**, 17266–17276.

Gernert KM, Richardson JS and Richardson DC (1993) Structural characteristics of FELIX, a designed protein. *Protein Engineering* **6**, S114–S114.

Gernert KM, Surles MC, Labean TH, Richardson JS and Richardson DC (1995) The Alacoil: a very tight, antiparallel coiled-coil of helices. *Protein Science* **4**, 2252–2260.

Ghadiri MR and Case MA (1993) *De-novo* design of a novel heteronuclear 3-helix bundle metalloprotein. *Angewandte Chemie* **32**, 1594–1597.

Ghadiri MR and Choi C (1990) Secondary structure nucleation in peptides. Transition metal ion stabilized α-helices. *Journal of the American Chemical Society* **112**, 1630–1632.

Ghirlanda G, Lear JD, Ogihara NL, Eisenberg D and DeGrado WF (2002) A hierarchic approach to the design of hexameric helical barrels. *Journal of Molecular Biology* **319**, 243–253.

Gonen S, DiMaio F, Gonen T and Baker D (2015) Design of ordered two-dimensional arrays mediated by noncovalent protein–protein interfaces. *Science* **348**, 1365–1368.

Goparaju G, Fry BA, Chobot SE, Wiedman G, Moser CC, Leslie Dutton P and Discher BM (2016) First principles design of a core bioenergetic transmembrane electron-transfer protein. *Biochimica et Biophysica Acta* **1857**, 503–512.

Goraj K, Renard A and Martial JA (1990) Synthesis, purification and initial structure characterization of octarellin, a *de novo* polypeptide modelled on the alpha/beta barrel proteins. *Protein Engineering* **3**, 259–266.

Gordon DB, Hom GK, Mayo SL and Pierce NA (2003) Exact rotamer optimization for protein design. *Journal of Computational Chemistry* **24**, 232–243.

Gradišar H, Božič S, Doles T, Vengust D, Hafner-Bratkovič I, Mertelj A, Webb B, Šali A, Klavžar S and Jerala R (2013) Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nature Chemical Biology* **9**, 362.

Gratkowski H, Lear JD and DeGrado WF (2001) Polar sidechains drive the association of model, transmembrane peptides. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 880–885.

Grayson KJ and Anderson JR (2018) The ascent of man(made oxidoreductases). *Current Opinion in Structural Biology* **51**, 149–155.

Grigoryan G and DeGrado WF (2011) Probing designability via a generalized model of helical bundle geometry. *Journal of Molecular Biology* **405**, 1079–1100.

Grigoryan G, Reinke AW and Keating AE (2009) Design of protein-interaction specificity gives selective bZIP-binding peptides. *Nature* **458**, 859–864.

Grigoryan G, Kim YH, Acharya R, Axelrod K, Jain RM, Willis L, Drndic M, Kikkawa JM and DeGrado WF (2011) Computational design of virus-like protein assemblies on carbon nanotube surfaces. *Science* **332**, 1071–1076.

Guo R, Gaffney K, Yang Z, Kim M, Sungsuwan S, Huang X, Hubbell WL and Hong H (2016) Steric trapping reveals a cooperativity network in the intramembrane protease GlpG. *Nature Chemical Biology* **12**, 353–360.

Gurezka R, Laage R, Brosig B and Langosch D (1999) A heptad motif of leucine residues found in membrane proteins can drive self-assembly of artificial transmembrane segments. *Journal of Biological Chemistry* **274**, 9265–9270.

Gutte B, Däumingen M and Wittschieber E (1979) Design, synthesis and characterization of a 34-residue polypeptide that interacts with nucleic acids. *Nature* **281**, 650–655.

Handel T and DeGrado WF (1990) *De novo* design of a Zn^{2+}-binding protein. *Journal of the American Chemical Society* **112**, 6710–6711.

Handel TM, Williams SA and DeGrado WF (1993) Metal ion-dependent modulation of the dynamics of a designed protein. *Science* **261**, 879–885.

Harbury PB, Zhang T, Kim PS and Alber T (1993) A switch between two-, three-, and four-stranded coiled coils. *Science* **262**, 1401–1407.

Harbury PB, Kim PS and Alber T (1994) Crystal structure of an isoleucine-zipper trimer. *Nature* **371**, 80–83.

Harbury PB, Tidor B and Kim PS (1995) Repacking protein cores with backbone freedom: structure prediction for coiled coils. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 8408–8412.

Harbury PB, Plecs JJ, Tidor B, Alber T and Kim PS (1998) High-resolution protein design with backbone freedom. *Science* **282**, 1462–1467.

Hecht MH (1994) *De novo* design of beta-sheet proteins. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 8729–8730.

Hecht MH, Richardson JS, Richardson DC and Ogden RC (1990) *De novo* design, expression and characterization of felix: a four-helix bundle protein of native-like sequence. *Science* **249**, 884–891.

Hecht MH, Das A, Go A, Bradley LH and Wei Y (2004) *De novo* proteins from designed combinatorial libraries. *Protein Science* **13**, 1711–1723.

Hecht MH, Zarzhitsky S, Karas C and Chari S (2018) Are natural proteins special? Can we do that? *Current Opinion in Structural Biology* **48**, 124–132.

Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, Nilsson I, White SH and von Heijne G (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377–381.

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

Hill RB and DeGrado WF (1998) Solution structure of alpha-2-D, a nativelike *de novo* designed protein. *Journal of the American Chemical Society* **120**, 1138–1145.

Hill RB and DeGrado WF (2000) A polar, solvent-exposed residue can be essential for native protein structure. *Structure: Folding and Design* **8**, 471–479.

Hill RB, Hong J-K and DeGrado WF (1999) Hydrogen bonding cluster can specify the unique conformation of a protein. *Journal of the American Chemical Society* **122**, 746–747.

Hill RB, Raleigh DP, Lombardi A and DeGrado WF (2000) *De novo* design of helical bundles as models for understanding protein folding and function. *Accounts of Chemical Research* **33**, 745–754.

Ho SP and DeGrado WF (1987) Design of a 4-helix bundle protein: synthesis of peptides which self-associate into a helical protein. *Journal of the American Chemical Society* **109**, 6751–6758.

Hong H (2014) Toward understanding driving forces in membrane protein folding. *Archives of Biochemistry and Biophysics* **564**, 297–313.

Houbrechts A, Moreau B, Abagyan R, Mainfroid V, Preaux G, Lamproye A, Poncin A, Goormaghtigh E, Ruysschaert JM, Martial JA, Goraj K.*et al.* (1995) Second-generation octarellins: two new *de novo* (beta/alpha)8 polypeptides designed for investigating the influence of beta-residue packing on the alpha/beta-barrel structure stability. *Protein Engineering* **8**, 249–259.

Hsia Y, Bale JB, Gonen S, Shi D, Sheffler W, Fong KK, Nattermann U, Xu C, Huang PS, Ravichandran R, Yi S, Davis TN, Gonen T, King NP and Baker D (2016) Design of a hyperstable 60-subunit protein dodecahedron [corrected]. *Nature* **535**, 136–139.

Huang SS, Koder RL, Lewis M, Wand AJ and Dutton PL (2004) The HP-1 maquette: from an apoprotein structure to a structured hemoprotein designed to promote redox-coupled proton exchange. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 5536–5541.

Huang PS, Ban YE, Richter F, Andre I, Vernon R, Schief WR and Baker D (2011) RosettaRemodel: a generalized framework for flexible backbone protein design. *PLoS One* **6**, e24109.

Huang PS, Oberdorfer G, Xu C, Pei XY, Nannenga BL, Rogers JM, DiMaio F, Gonen T, Luisi B and Baker D (2014) High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481–485.

Huang PS, Boyken SE and Baker D (2016*a*) The coming of age of *de novo* protein design. *Nature* **537**, 320–327.

Huang PS, Feldmeier K, Parmeggiani F, Velasco DAF, Hocker B and Baker D (2016*b*) *De novo* design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nature Chemical Biology* **12**, 29–34.

Hughes SA, Wang F, Wang S, Kreutzberger MAB, Osinski T, Orlova A, Wall JS, Zuo X, Egelman EH and Conticello VP (2019) Ambidextrous helical nanotubes from self-assembly of designed helical hairpin motifs. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 14456-14464.

Iranzo O, Chakraborty S, Hemmingsen L and Pecoraro VL (2011) Controlling and fine tuning the physical properties of two identical metal coordination sites in *de novo* designed three stranded coiled coil peptides. *Journal of the American Chemical Society* **133**, 239–251.

Janin J, Wodak S, Levitt M and Maigret B (1978) Conformation of amino acid side-chains in proteins. *Journal of Molecular Biology* **125**, 357??386.

Jasniewski AJ and Que L (2018) Dioxygen activation by nonheme diiron enzymes: diverse dioxygen adducts, high-valent intermediates, and related model complexes. *Chemical Reviews* **118**, 2554–2592.

Joh NH, Oberai A, Yang D, Whitelegge JP and Bowie JU (2009) Similar energetic contributions of packing in the core of membrane and water-soluble proteins. *Journal of the American Chemical Society* **131**, 10846–10847.

Joh NH, Wang T, Bhate MP, Acharya R, Wu Y, Grabe M, Hong M, Grigoryan G and DeGrado WF (2014) *De novo* design of a transmembrane Zn^{2+}-transporting four-helix bundle. *Science* **346**, 1520–1524.

Joh NH, Grigoryan G, Wu Y and DeGrado WF (2017) Design of self-assembling transmembrane helical bundles to elucidate principles required for membrane protein folding and ion transport. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **372**, 20160214.

Johnson EC, Lazar GA, Desjarlais JR and Handel TM (1999) Solution structure and dynamics of a designed hydrophobic core variant of ubiquitin. *Structure: Folding and Design* **7**, 967–976.

Johnson RM, Heslop CL and Deber CM (2004) Hydrophobic helical hairpins: design and packing interactions in membrane environments. *Biochemistry* **43**, 14361–14369.

Jones DT (1994) *De-novo* protein design using pairwise potentials and a genetic algorithm. *Protein Science* **3**, 567–574.

Jumper JM, Faruk NF, Freed KF and Sosnick TR (2018) Trajectory-based training enables protein simulations with accurate folding and Boltzmann ensembles in cpu-hours. *PLoS Computational Biology* **14**, e1006578.

Kajander T, Cortajarena AL and Regan L (2006) Consensus design as a tool for engineering repeat proteins. *Methods in Molecular Biology* **340**, 151–170.

Kajava AV (2012) Tandem repeats in proteins: from sequence to structure. *Journal of Structural Biology* **179**, 279–288.

Kalb SR, Garcia-Rodriguez C, Lou J, Baudys J, Smith TJ, Marks JD, Smith LA, Pirkle JL and Barr JR (2010) Extraction of BoNT/A, /B, /E, and /F with a single, high affinity monoclonal antibody for detection of botulinum neurotoxin by Endopep-MS. *PLoS One* **5**, e12237.

Kamtekar S, Schiffer JM, Xiong H, Babik JM and Hecht MH (1993) Protein design by binary patterning of polar and nonpolar amino acids. *Science* **262**, 1680–1685.

Kang ES, Kim YT, Ko YS, Kim NH, Cho G, Huh YH, Kim JH, Nam J, Thach TT, Youn D, Kim YD, Yun WS, DeGrado WF, Kim SY, Hammond PT, Lee J, Kwon YU, Ha DH and Kim YH (2018) Peptide-programmable nanoparticle superstructures with tailored electrocatalytic activity. *ACS Nano*, **12**, 6554-6562.

Kaplan J and DeGrado WF (2004) *De novo* design of catalytic proteins. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 11566–11570.

Katz B (1969) *The Release of Neural Transmitter Substances*. Liverpool University Press, Liverpool.

Kennedy ML and Gibney BR (2001) Metalloprotein and redox protein design. *Current Opinion in Structural Biology* **11**, 485–490.

Kim KH, Ko DK, Kim YT, Kim NH, Paul J, Zhang SQ, Murray CB, Acharya R, DeGrado WF, Kim YH and Grigoryan G (2016) Protein-directed self-assembly of a fullerene crystal. *Nature Communications* **7**, 11429.

King NP, Sheffler W, Sawaya MR, Vollmar BS, Sumida JP, Andre I, Gonen T, Yeates TO and Baker D (2012) Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336**, 1171–1174.

Kirrbach J, Krugliak M, Ried CL, Pagel P, Arkin IT and Langosch D (2013) Self-interaction of transmembrane helices representing pre-clusters from the human single-span membrane proteins. *Bioinformatics (Oxford, England)* **29**, 1623–1630.

Kiyokawa T, Kanaori K, Tajima K, Koike M, Mizuno T, Oku JI and Tanaka T (2004) Binding of Cu(II) or Zn(II) in a *de novo* designed triple-stranded alpha-helical coiled-coil toward a prototype for a metalloenzyme. *Journal of Peptide Research* **63**, 347–353.

Kobayashi N and Arai R (2017) Design and construction of self-assembling supramolecular protein complexes using artificial and fusion proteins as nanoscale building blocks. *Current Opinion in Biotechnology* **46**, 57–65.

Kobe B and Kajava AV (2000) When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends in Biochemical Sciences* **25**, 509–515.

Kodali G, Mancini JA, Solomon LA, Episova TV, Roach N, Hobbs CJ, Wagner P, Mass OA, Aravindu K, Barnsley JE, Gordon KC, Officer DL, Dutton PL and Moser CC (2017) Design and engineering of water-soluble light-harvesting protein maquettes. *Chemical Science* **8**, 316–324.

Koder RL, Anderson JL, Solomon LA, Reddy KS, Moser CC and Dutton PL (2009) Design and engineering of an O(2) transport protein. *Nature* **458**, 305–309.

Koebke KJ, Ruckthong L, Meagher JL, Mathieu E, Harland J, Deb A, Lehnert N, Policar C, Tard C, Penner-Hahn JE, Stuckey JA and

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

Pecoraro VL (2018) Clarifying the copper coordination environment in a *de novo* designed red copper protein. *Inorganic Chemistry* **57**, 12291–12302.

Koehler Leman J, Mueller BK and Gray JJ (2017) Expanding the toolkit for membrane protein modeling in Rosetta. *Bioinformatics (Oxford, England)* **33**, 754–756.

Korendovych IV, Senes A, Kim YH, Lear JD, Fry HC, Therien MJ, Blasie JK, Walker FA and DeGrado WF (2010) *De novo* design and molecular assembly of a transmembrane diporphyrin-binding protein complex. *Journal of the American Chemical Society* **132**, 15516–15518.

Kortemme T, Ramirez-Alvarado M and Serrano L (1998) Design of a 20-amino acid, three-stranded beta-sheet protein. *Science* **281**, 253–256.

Krantz BA and Sosnick TR (2001) Engineered metal binding sites map the heterogeneous folding landscape of a coiled coil. *Natural Structural Biology* **8**, 1042–1047.

Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL and Baker D (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368.

Kumar M, Ing NL, Narang V, Wijerathne NK, Hochbaum AI and Ulijn RV (2018) Amino-acid-encoded biocatalytic self-assembly enables the formation of transient conducting nanostructures. *Nature Chemistry* **10**, 696–703.

Lahr SJ, Engel DE, Stayrook SE, Maglio O, North B, Geremia S, Lombardi A and DeGrado WF (2005) Analysis and design of turns in alpha-helical hairpins. *Journal of Molecular Biology* **346**, 1441–1454.

Lanci CJ, MacDermaid CM, Kang SG, Acharya R, North B, Yang X, Qiu XJ, DeGrado WF and Saven JG (2012) Computational design of a protein crystal. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 7304–7309.

Langosch D and Arkin IT (2009) Interaction and conformational dynamics of membrane-spanning protein helices. *Protein Science* **18**, 1343–1358.

Langosch D and Heringa J (1998) Interaction of transmembrane helices by a knobs-into-holes packing characteristic of soluble coiled coils. *Proteins* **31**, 150–159.

Langosch D, Brosig B, Kolmar H and Fritz HJ (1996) Dimerisation of the glycophorin A transmembrane segment in membranes probed with the ToxR transcription activator. *Journal of Molecular Biology* **263**, 525–530.

Lapenta F, Aupič J, Strmšek Ž and Jerala R (2018) Coiled coil protein origami: from modular design principles towards biotechnological applications. *Chem. Soc. Rev* **47**(10), 3530–3542.

Lasters I, Wodak SJ, Alard P and van Cutsem E (1988) Structural principles of parallel beta-barrels in proteins. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 3338–3342.

Lasters I, De Maeyer M and Desmet J (1995) Enhanced dead-end elimination in the search for the global minimum energy conformation of a collection of protein side chains. *Protein Engineering* **8**, 815–822.

Lau SYM, Taneja AK and Hodges RS (1984) Synthesis of a model protein of defined secondary and quaternary structure. *Journal of Biological Chemistry* **259**, 13253–13261.

Lazar GA, Desjarlais JR and Handel TM (1997) *De novo* design of the hydrophobic core of ubiquitin. *Protein Science* **6**, 1167–1178.

Lear JD, Wasserman ZR and DeGrado WF (1988) Synthetic amphiphilic peptide models for protein ion channels. *Science* **240**, 1177–1181.

Lear JD, Gratkowski H, Adamian L, Liang J and DeGrado WF (2003) Position-dependence of stabilizing polar interactions of asparagine in transmembrane helical bundles. *Biochemistry* **42**, 6400–6407.

Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F, Ban YE, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popović Z, Havranek JJ, Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D and Bradley P (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods in Enzymology* **487**, 545–574.

Lee M, Wang T, Makhlynets OV, Wu Y, Polizzi NF, Wu H, Gosavi PM, Stohr J, Korendovych IV, DeGrado WF and Hong M (2017) Zinc-binding structure of a catalytic amyloid from solid-state NMR. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 6191–6196.

Levinthal C (1969) How to fold graciously. *Mossbauer Spectroscopy in Biological Systems* **67**, 22–24.

Lichtenstein BR, Farid TA, Kodali G, Solomon LA, Anderson JL, Sheehan MM, Ennist NM, Fry BA, Chobot SE, Bialas C, Mancini JA, Armstrong CT, Zhao Z, Esipova TV, Snell D, Vinogradov SA, Discher BM, Moser CC and Dutton PL (2012) Engineering oxidoreductases: maquette proteins designed from scratch. *Biochemical Society Transactions* **40**, 561–566.

Lim A, Saderholm MJ, Makhov AM, Kroll M, Yan Y, Perera L, Griffith JD and Erickson BW (1998) Engineering of betabellin-15D: a 64 residue beta sheet protein that forms long narrow multimeric fibrils. *Protein Science* **7**, 1545–1554.

Lin YR, Koga N, Tatsumi-Koga R, Liu G, Clouser AF, Montelione GT and Baker D (2015) Control over overall shape and size in *de novo* designed proteins. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E5478–E5485.

Liu F, Dumont C, Zhu Y, DeGrado WF, Gai F and Gruebele M (2009) A one-dimensional free energy surface does not account for two-probe folding kinetics of protein alpha(3)D. *Journal of Chemical Physics* **130**, 061101.

Ljubetič A, Lapenta F, Gradišar H, Drobnak I, Aupič J, Strmšek Ž, Lainšček D, Hafner-Bratkovič I, Majerle A, Krivec N, Benčina M, Pisanski T, Veličković TĆ, Round A, Carazo JM, Melero R and Jerala R (2017) Design of coiled-coil protein-origami cages that self-assemble *in vitro* and *in vivo*. *Nature Biotechnology* **35**, 1094–1101.

Lombardi A, Marasco D, Maglio O, Di Costanzo L, Nastri F and Pavone V (2000*a*) Miniaturized metalloproteins: application to iron-sulfur proteins. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 11922–11927.

Lombardi A, Summa CM, Geremia S, Randaccio L, Pavone V and DeGrado WF (2000*b*) Retrostructural analysis of metalloproteins: application to the design of a minimal model for diiron proteins. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 6298–6305.

Lombardi A, Nastri F and Pavone V (2001) Peptide-based heme-protein models. *Chemical Reviews* **101**, 3165–3189.

Lombardi A, Pirro F, Maglio O, Chino M and DeGrado WF (2019) *De novo* design of four-helix bundle metalloproteins: one scaffold, diverse reactivities. *Accounts of Chemical Research* **52**, 1148–1159.

Lomize AL, Lomize MA, Krolicki SR and Pogozheva ID (2017) Membranome: a database for proteome-wide analysis of single-pass membrane proteins. *Nucleic Acids Research* **45**, D250–D255.

Lu P, Min D, DiMaio F, Wei KY, Vahey MD, Boyken SE, Chen Z, Fallas JA, Ueda G, Sheffler W, Mulligan VK, Xu W, Bowie JU and Baker D (2018) Accurate computational design of multipass transmembrane proteins. *Science* **359**, 1042–1046.

MacKenzie KR and Fleming KG (2008) Association energetics of membrane spanning alpha-helices. *Current Opinion in Structural Biology* **18**, 412–419.

Mackenzie CO and Grigoryan G (2017) Protein structural motifs in prediction and design. *Current Opinion in Structural Biology* **44**, 161–167.

MacKenzie KR, Prestegard JH and Engelman DM (1997) A transmembrane helix dimer: structure and implications. *Science* **276**, 131–133.

Mackenzie CO, Zhou J and Grigoryan G (2016) Tertiary alphabet for the observable protein structural universe. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E7438–E7447.

Maglio O, Nastri F, Pavone V, Lombardi A and DeGrado WF (2003) Preorganization of molecular binding sites in designed diiron proteins. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 3772–3777.

Makhlynets OV and Korendovych IV (2017) Finding a silver bullet in a stack of proteins. *Biochemistry* **56**, 6627–6628.

Marcos E, Basanta B, Chidyausiku TM, Tang Y, Oberdorfer G, Liu G, Swapna GV, Guan R, Silva DA, Dou J, Pereira JH, Xiao R, Sankaran B, Zwart PH, Montelione GT and Baker D (2017) Principles for designing proteins with cavities formed by curved beta sheets. *Science* **355**, 201–206.

Marcos E, Chidyausiku TM, McShan AC, Evangelidis T, Nerli S, Carter L, Nivon LG, Davis A, Oberdorfer G, Tripsianes K, Sgourakis NG and Baker D (2018) *De novo* design of a non-local beta-sheet protein with high stability and accuracy. *Nature Structural & Molecular Biology* **25**, 1028–1034.

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

Marsh EN and DeGrado WF (2002) Noncovalent self-assembly of a heterotetrameric diiron protein. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 5150–5154.

Marsh ENG and Waugh MW (2013) Aldehyde decarbonylases: enigmatic enzymes of hydrocarbon biosynthesis. *ACS Catalysis* **3**, 2515–2521.

Maruyama Y and Mitsutake A (2017) Stability of unfolded and folded protein structures using a 3D-RISM with the RMDFT. *Journal of Physical Chemistry B* **121**, 9881–9885.

McGregor MJ, Islam SA and Sternberg MJ (1987) Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *Journal of Molecular Biology* **198**, 295–310.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH and Teller E (1953) Equation of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1092.

Miller S, Janin J, Lesk AM and Chothia C (1987) Interior and surface of monomeric proteins. *Journal of Molecular Biology* **196**, 641–656.

Mocny CS and Pecoraro VL (2015) *De novo* protein design as a methodology for synthetic bioinorganic chemistry. *Accounts of Chemical Research* **48**, 2388–2396.

Morein S, Koeppe IR, Lindblom G, de Kruijff B and Killian JA (2000) The effect of peptide/lipid hydrophobic mismatch on the phase behavior of model membranes mimicking the lipid composition in *Escherichia coli* membranes. *Biophysical Journal* **78**, 2475–2485.

Moser R, Thomas RM and Gutte B (1983) An artificial crystalline DDT-binding peptide. *FEBS Letters* **157**, 247–251.

Mravic M, Hu H, Lu Z, Bennett JS, Sanders CR, Orr AW and DeGrado WF (2018) *De novo* designed transmembrane peptides activating the alpha5-beta1 integrin. *Protein Engineering, Design & Selection* **31**, 181–190.

Mravic M, Thomaston JL, Tucker M, Solomon PE, Liu L and DeGrado WF (2019) Packing of apolar side chains enables accurate design of highly stable membrane proteins. *Science* **363**, 1418–1423.

Mueller BK, Subramaniam S and Senes A (2014) A frequent, GxxxG-mediated, transmembrane association motif is optimized for the formation of interhelical Calpha-H hydrogen bonds. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E888–E895.

Murase S, Ishino S, Ishino Y and Tanaka T (2012) Control of enzyme reaction by a designed metal-ion-dependent alpha-helical coiled-coil protein. *Journal of Biological Inorganic Chemistry* **17**, 791–799.

Mustata GM, Kim YH, Zhang J, DeGrado WF, Grigoryan G and Wanunu M (2016) Graphene symmetry amplified by designed peptide self-assembly. *Biophysical Journal* **110**, 2507–2516.

Nagy-Smith K, Moore E, Schneider J and Tycko R (2015) Molecular structure of monomorphic peptide fibrils within a kinetically trapped hydrogel network. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 9816–9821.

Nambiar M, Wang LS, Rotello V and Chmielewski J (2018) Reversible hierarchical assembly of trimeric coiled-coil peptides into banded nano- and microstructures. *Journal of the American Chemical Society* **140**, 13028–13033.

Nanda V, Rosenblatt MM, Osyczka A, Kono H, Getahun Z, Dutton PL, Saven JG and Degrado WF (2005) *De novo* design of a redox-active minimal rubredoxin mimic. *Journal of the American Chemical Society* **127**, 5804–5805.

Nguyen TK and Ueno T (2018) Engineering of protein assemblies within cells. *Current Opinion in Structural Biology* **51**, 1–8.

Nguyen TH, Liu Z and Moore PB (2013) Molecular dynamics simulations of homo-oligomeric bundles embedded within a lipid bilayer. *Biophysical Journal* **105**, 1569–1580.

Norn CH and Andre I (2016) Computational design of protein self-assembly. *Current Opinion in Structural Biology* **39**, 39–45.

North B, Summa CM, Ghirlanda G and DeGrado WF (2001) D (n)-symmetrical tertiary templates for the design of tubular proteins. *Journal of Molecular Biology* **311**, 1081–1090.

Oberai A, Joh NH, Pettit FK and Bowie JU (2009) Structural imperatives impose diverse evolutionary constraints on helical membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 17747–17750.

Offer G, Hicks MR and Woolfson DN (2002) Generalized Crick equations for modeling noncanonical coiled coils. *Journal of Structural Biology* **137**, 41–53.

Ogihara NL, Weiss MS, DeGrado WF and Eisenberg D (1997) The crystal structure of the designed trimeric coiled coil coil-V_aL_d: implications for engineering crystals and supramolecular assemblies. *Protein Science* **6**, 80–88.

Ogihara NL, Ghirlanda G, Bryson JW, Gingery M, DeGrado WF and Eisenberg D (2001) Design of three-dimensional domain-swapped dimers and fibrous oligomers. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 1404–1409.

O'Neil KT and DeGrado WF (1990) A thermodynamic scale for the helix forming tendencies of the commonly occurring amino acids. *Science* **250**, 646–651.

Osterhout JJ, Handel T, Na G, Toumadje A, Long RC, Connolly PJ, Hoch JC, Johnson WC, Live D and DeGrado WF (1992) Characterization of the structural properties of a_1b, a peptide designed to form a four-helix bundle. *Journal of the American Chemical Society*, **114**, 331–337.

Osterman DG and Kaiser ET (1985) Design and characterization of peptides with amphiphilic beta-strand structures. *Journal of Cellular Biochemistry* **29**, 57–72.

Padilla JE, Colovos C and Yeates TO (2001) Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 2217–2221.

Pandya MJ, Spooner GM, Sunde M, Thorpe JR, Rodger A and Woolfson DN (2000) Sticky-end assembly of a designed peptide fiber provides insight into protein fibrillogenesis. *Biochemistry* **39**, 8728–8734.

Park S, Xu Y, Stowell XF, Gai F, Saven JG and Boder ET (2006) Limitations of yeast surface display in engineering proteins of high thermostability. *Protein Engineering, Design & Selection* **19**, 211–217.

Park WM, Bedewy M, Berggren KK and Keating AE (2017) Modular assembly of a protein nanotriangle using orthogonally interacting coiled coils. *Scientific Reports* **7**, 10577.

Partridge AW, Therien AG and Deber CM (2004) Missense mutations in transmembrane domains of proteins: phenotypic propensity of polar residues for human disease. *Proteins* **54**, 648–656.

Pasternak A, Kaplan J, Lear JD and Degrado WF (2001) Proton and metal ion-dependent assembly of a model diiron protein. *Protein Science* **10**, 958–969.

Patterson WR, Anderson DH, DeGrado WF, Cascio D and Eisenberg D (1999) Centrosymmetric bilayers in the 0.75 A resolution structure of a designed alpha-helical peptide, D,L-alpha-1. *Protein Science* **8**, 1410–1422.

Pauling L and Corey RB (1951) Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. *Proceedings of the National Academy of Sciences of the United States of America* **37**, 729–740.

Pellach M, Mondal S, Harlos K, Mance D, Baldus M, Gazit E and Shimon LJ (2017) A two-tailed phosphopeptide crystallizes to form a lamellar structure. *Angewandte Chemie (International Edition)* **56**, 3252–3255.

Plegaria JS and Pecoraro VL (2016) *De novo* design of metalloproteins and metalloenzymes in a three-helix bundle. *Methods in Molecular Biology* **1414**, 187–196.

Pluckthun A (2015) Designed ankyrin repeat proteins (DARPins): binding proteins for research, diagnostics, and therapy. *Annual Review of Pharmacology and Toxicology* **55**, 489–511.

Polizzi NF, Wu Y, Lemmin T, Maxwell AM, Zhang SQ, Rawson J, Beratan DN, Therien MJ and DeGrado WF (2017) *De novo* design of a hyperstable non-natural protein–ligand complex with sub-A accuracy. *Nature Chemistry* **9**, 1157–1164.

Ponder JW and Richards FM (1987) Tertiary templates for proteins use of packing criteria in the enumeration of allowed sequences for different structural classes. *Journal of Molecular Biology* **193**, 775–791.

Presnell SR and Cohen FE (1989) Topological distribution of four-a-helix bundles. *Proceedings of the National Academy of Sciences of the USA* **86**, 6592–6596.

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

Prive GG, Anderson DH, Wesson L, Cascio D and Eisenberg D (1999) Packed protein bilayers in the 0.90 A resolution structure of a designed alpha helical bundle. *Protein Science* **8**, 1400–1409.

Quinn TP, Tweedy NB, Williams RW, Richardson JS and Richardson DC (1994) Betadoublet: *de novo* design, synthesis, and characterization of a beta-sandwich protein. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 8747–8751.

Ramisch S, Weininger U, Martinsson J, Akke M and Andre I (2014) Computational design of a leucine-rich repeat protein with a predefined geometry. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 17875–17880.

Randa HS, Forrest LR, Voth GA and Sansom MS (1999) Molecular dynamics of synthetic leucine-serine ion channels in a phospholipid membrane. *Biophysical Journal* **77**, 2400–2410.

Razkin J, Nilsson H and Baltzer L (2007) Catalysis of the cleavage of uridine 3′−2,2,2-trichloroethylphosphate by a designed helix–loop–helix motif peptide. *Journal of the American Chemical Society* **129**, 14752–14758.

Razkin J, Lindgren J, Nilsson H and Baltzer L (2008) Enhanced complexity and catalytic efficiency in the hydrolysis of phosphate diesters by rationally designed helix–loop–helix motifs. *ChemBioChem* **9**, 1975–1984.

Reedy CJ and Gibney BR (2004) Heme protein assemblies. *Chemical Reviews* **104**, 617–649.

Regan L and Clarke ND (1990) A tetrahedral zinc(II)-binding site introduced into a designed protein. *Biochemistry* **29**, 10878–10883.

Regan L and DeGrado WF (1988) Characterization of a helical protein designed from first principles. *Science* **241**, 976–978.

Regan L, Rockwell A, Wasserman Z and DeGrado W (1994) Disulfide cross-links to probe the structure and flexibility of a designed four-helix bundle protein. *Protein Science* **3**, 2419–2427.

Reig AJ, Pires MM, Snyder RA, Wu Y, Jo H, Kulp DW, Butch SE, Calhoun JR, Szyperski TG, Solomon EI and DeGrado WF (2012) Alteration of the oxygen-dependent reactivity of *de novo* Due Ferri proteins. *Nature Chemistry* **4**, 900–906.

Ren J, Lew S, Wang J and London E (1999) Control of the transmembrane orientation and interhelical interactions within membranes by hydrophobic helix length. *Biochemistry* **38**, 5905–5912.

Richards FM (1977) Areas, volumes, packing and protein structure. *Annual Review of Biophysics and Bioengineering* **6**, 151–176.

Richardson JS and Richardson DC (1989) The *de novo* design of protein structures. *Trends in Biochemical Sciences* **14**, 304–309.

Richardson JS and Richardson DC (2002) Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 2754–2759.

Robertson DE, Farid RS, Moser CC, Urbauer JL, Mulholland SE, Pidikiti R, Lear JD, Wand AJ, DeGrado WF and Dutton PL (1994) Design and synthesis of multi-haem proteins. *Nature* **368**, 425–432.

Ruan F, Chen Y and Hopkins PB (1990) Metal ion enhanced helicity in synthetic peptides containing unnatural, metal-ligating residues. *Journal of the American Chemical Society* **112**, 9403–9404.

Rufo CM, Moroz YS, Moroz OV, Stohr J, Smith TA, Hu X, DeGrado WF and Korendovych IV (2014) Short peptides self-assemble to produce catalytic amyloids. *Nature Chemistry* **6**, 303–309.

Salemme FR (1983) Structural properties of protein beta-sheets. *Progress in Biophysics & Molecular Biology* **42**, 95–133.

Salgado EN, Faraone-Mennella J and Tezcan FA (2007) Controlling protein–protein interactions through metal coordination: assembly of a 16-helix bundle protein. *Journal of the American Chemical Society* **129**, 13374–13375.

Salgado EN, Radford RJ and Tezcan FA (2010) Metal-directed protein self-assembly. *Accounts of Chemical Research* **43**, 661–672.

Schafmeister CE, LaPorte SL, Miercke LJ and Stroud RM (1997) A designed four helix bundle protein with native-like structure. *Nature Structural Biology* **4**, 1039–1046.

Schenck HL and Gellman SH (1998) Use of a designed triple-stranded antiparallel beta-sheet to probe beta-sheet cooperativity in aqueous solution. *Journal of the American Chemical Society* **120**, 4869–4870.

Schlebach JP and Sanders CR (2015) The safety dance: biophysics of membrane protein folding and misfolding in a cellular context. *Quarterly Reviews of Biophysics* **48**, 1–34.

Schneider JP, Pochan DJ, Ozbas B, Rajagopal K, Pakstis L and Kretsinger J (2002) Responsive hydrogels from the intramolecular folding and self-assembly of a designed peptide. *Journal of the American Chemical Society* **124**, 15030–15037.

Schramm CA, Hannigan BT, Donald JE, Keasar C, Saven JG, Degrado WF and Samish I (2012) Knowledge-based potential for positioning membrane-associated structures and assessing residue-specific energetic contributions. *Structure* **20**, 924–935.

Sciore A, Su M, Koldewey P, Eschweiler JD, Diffley KA, Linhares BM, Ruotolo BT, Bardwell JCA, Skiniotis G and Marsh ENG (2016) Flexible, symmetry-directed approach to assembling protein cages. *Proceedings of the National Academy of Sciences* **113**, 8681.

Senes A, Gerstein M and Engelman DM (2000) Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. *Journal of Molecular Biology* **296**, 921–936.

Senes A, Ubarretxena-Belandia I and Engelman DM (2001) The Calpha---H…O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 9056–9061.

Senes A, Chadi DC, Law PB, Walters RF, Nanda V and Degrado WF (2007) E(z), a depth-dependent potential for assessing the energies of insertion of amino acid side-chains into membranes: derivation and applications to determining the orientation of transmembrane and interfacial helices. *Journal of Molecular Biology* **366**, 436–448.

Shandler SJ, Shapovalov MV, Dunbrack Jr RL and DeGrado WF (2010) Development of a rotamer library for use in beta-peptide foldamer computational design. *Journal of the American Chemical Society* **132**, 7312–7320.

Shao Q (2014) Probing sequence dependence of folding pathway of alpha-helix bundle proteins through free energy landscape analysis. *Journal of Physical Chemistry B* **118**, 5891–5900.

Sharman GJ and Searle MS (1998) Cooperative interaction between the three strands of a designed antiparallel beta-sheet. *Journal of the American Chemical Society* **120**, 5291–5300.

Shen H, Fallas JA, Lynch E, Sheffler W, Parry B, Jannetty N, Decarreau J, Wagenbach M, Vicente JJ, Chen J, Wang L, Dowling Q, Oberdorfer G, Stewart L, Wordeman L, De Yoreo J, Jacobs-Wagner C, Kollman J and Baker D (2018) *De novo* design of self-assembling helical protein filaments. *Science* **362**, 705.

Shigemitsu H and Hamachi I (2017) Design strategies of stimuli-responsive supramolecular hydrogels relying on structural analyses and cell-mimicking approaches. *Accounts of Chemical Research* **50**, 740–750.

Shoichet BK, Baase WA, Kuroki R and Matthews BW (1995) A relationship between protein stability and protein function. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 452–456.

Signarvic RS and DeGrado WF (2009) Metal-binding dependent disruption of membranes by designed helices. *Journal of the American Chemical Society* **131**, 3377-3384.

Snyder RA, Betzu J, Butch SE, Reig AJ, DeGrado WF and Solomon EI (2015) Systematic perturbations of binuclear non-heme iron sites: structure and dioxygen reactivity of *de novo* due Ferri proteins. *Biochemistry* **54**, 4637–4651.

Song WJ and Tezcan FA (2014) A designed supramolecular protein assembly with *in vivo* enzymatic activity. *Science* **346**, 1525–1528.

Struthers MD, Cheng RP and Imperiali B (1996*a*) Design of a monomeric 23-residue polypeptide with defined tertiary structure. *Science* **271**, 342–345.

Struthers MD, Cheng RP and Imperiali B (1996*b*) Economy in protein design: evolution of a metal-independent beta beta alpha motif based on the zinc finger domains. *Journal of the American Chemical Society* **118**, 3073–3081.

Studer S, Hansen DA, Pianowski ZL, Mittl PRE, Debon A, Guffy SL, Der BS, Kuhlman B and Hilvert D (2018) Evolution of a highly active and enantiospecific metalloenzyme from short peptides. *Science* **362**, 1285–1288.

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

Summa CM, Lombardi A, Lewis M and DeGrado WF (1999) Tertiary templates for the design of diiron proteins. *Current Opinion in Structural Biology* **9**, 500–508.

Summa CM, Rosenblatt MM, Hong JK, Lear JD and DeGrado WF (2002) Computational *de novo* design, and characterization of an A(2)B(2) diiron protein. *Journal of Molecular Biology* **321**, 923–938.

Suzuki K, Hiroaki H, Kohda D and Tanaka T (1998) An isoleucine zipper peptide forms a native-like triple stranded coiled coil in solution. *Protein Engineering* **11**, 1051–1055.

Tanaka R, Kimura H, Hayashi M, Fujiyoshi Y, Fukuhara K-I and Nakamura H (1994) Characteristics of a *de novo* designed protein. *Protein Science* **3**, 419–427.

Tanaka T, Mizuno T, Fukui S, Hiroaki H, Oku J, Kanaori K, Tajima K and Shirakawa M (2004) Two-metal ion, Ni(II) and Cu(II), binding alpha-helical coiled coil peptide. *Journal of the American Chemical Society* **126**, 14023–14028.

Tang J, Signarvic RS, DeGrado WF and Gai F (2007) Role of helix nucleation in the kinetics of binding of mastoparan X to phospholipid bilayers. *Biochemistry* **46**, 13856–13863.

Tatko CD, Nanda V, Lear JD and Degrado WF (2006) Polar networks control oligomeric assembly in membranes. *Journal of the American Chemical Society* **128**, 4170–4171.

Tayeb-Fligelman E, Tabachnikov O, Moshe A, Goldshmidt-Tran O, Sawaya MR, Coquelle N, Colletier JP and Landau M (2017) The cytotoxic Staphylococcus aureus PSMalpha3 reveals a cross-alpha amyloid-like fibril. *Science* **355**, 831–833.

Tebo AG and Pecoraro VL (2015) Artificial metalloenzymes derived from three-helix bundles. *Current Opinion in Chemical Biology* **25**, 65–70.

Tegoni M, Yu F, Bersellini M, Penner-Hahn JE and Pecoraro VL (2012) Designing a functional type 2 copper center that has nitrite reductase activity within alpha-helical coiled coils. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 21234–21239.

Thomson AR, Wood CW, Burton AJ, Bartlett GJ, Sessions RB, Brady RL and Woolfson DN (2014) Computational design of water-soluble alpha-helical barrels. *Science* **346**, 485–488.

Ulas G, Lemmin T, Wu Y, Gassner GT and DeGrado WF (2016) Designed metalloprotein stabilizes a semiquinone radical. *Nature Chemistry* **8**, 354–359.

Unson CG, Erickson BW, Richardson DC and Richardson JS (1984) *Federation Proceedings* **43**, A1837.

Unterreitmeier S, Fuchs A, Schaffler T, Heym RG, Frishman D and Langosch D (2007) Phenylalanine promotes interaction of transmembrane domains via GxxxG motifs. *Journal of Molecular Biology* **374**, 705–718.

Voet AR, Noguchi H, Addy C, Simoncini D, Terada D, Unzai S, Park SY, Zhang KY and Tame JR (2014) Computational design of a self-assembling symmetrical beta-propeller protein. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 15102–15107.

Wagner DE, Phillips CL, Ali WM, Nybakken GE, Crawford ED, Schwab AD, Smith WF and Fairman R (2005) Toward the development of peptide nanofilaments and nanoropes as smart materials. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 12656–12661.

Walder R, LeBlanc MA, Van Patten WJ, Edwards DT, Greenberg JA, Adhikari A, Okoniewski SR, Sullan RMA, Rabuka D, Sousa MC and Perkins TT (2017) Rapid characterization of a mechanically labile alpha-helical protein enabled by efficient site-specific bioconjugation. *Journal of the American Chemical Society* **139**, 9867–9875.

Walsh ST, Cheng H, Bryson JW, Roder H and DeGrado WF (1999) Solution structure and dynamics of a *de novo* designed three-helix bundle protein. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 5486–5491.

Walters RF and DeGrado WF (2006) Helix-packing motifs in membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 13658–13663.

Wang W, Liang AD and Lippard SJ (2015) Coupling oxygen consumption with hydrocarbon oxidation in bacterial multicomponent monooxygenases. *Accounts of Chemical Research* **48**, 2632–2639.

Watkins DW, Jenkins JMX, Grayson KJ, Wood N, Steventon JW, Le Vay KK, Goodwin MI, Mullen AS, Bailey HJ, Crump MP, MacMillan F, Mulholland AJ, Cameron G, Sessions RB, Mann S and Anderson JLR (2017) Construction and *in vivo* assembly of a catalytically proficient and hyperthermostable *de novo* enzyme. *Nature Communications* **8**, 358.

Webber MJ, Appel EA, Meijer EW and Langer R (2016) Supramolecular biomaterials. *Nature Materials* **15**, 13–26.

Weber PC and Salemme FR (1980) Structural and functional diversity in 4-a-helical proteins. *Nature* **287**, 82–84.

West MW, Wang W, Patterson J, Mancias JD, Beasley JR and Hecht MH (1999) *De novo* amyloid proteins from designed combinatorial libraries. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 11211–11216.

White SH and von Heijne G (2005) Do protein–lipid interactions determine the recognition of transmembrane helices at the ER translocon? *Biochemical Society Transactions* **33**(Pt 5), 1012–1015.

White SH and von Heijne G (2008) How translocons select transmembrane helices. *Annual Review of Biophysics* **37**, 23–42.

Whitley P, Nilsson I and von Heijne G (1994) *De novo* design of integral membrane proteins. *Natural Structural Biology* **1**, 858–862.

Willett P (1995) Genetic algorithms in molecular recognition and design. *Trends in Biotechnology* **13**, 516–521.

Wilmot CM and Thornton JM (1988) Analysis and prediction of the different types of beta-turn in proteins. *Journal of Molecular Biology* **203**, 221–232.

Wolynes PG (2015) Evolution, energy landscapes and the paradoxes of protein folding. *Biochimie* **119**, 218–230.

Woolfson DN, Bartlett GJ, Burton AJ, Heal JW, Niitsu A, Thomson AR and Wood CW (2015) *De novo* protein design: how do we expand into the universe of possible protein structures? *Current Opinion in Structural Biology* **33**, 16–26.

Xiong DP, Mao WZ and Gong HP (2017) Predicting the helix–helix interactions from correlated residue mutations. *Proteins: Structure, Function, and Bioinformatics* **85**, 2162–2169.

Xu C, Liu R, Mehta AK, Guerrero-Ferreira RC, Wright ER, Dunin-Horkawicz S, Morris K, Serpell LC, Zuo X, Wall JS and Conticello VP (2013) Rational design of helical nanotubes from self-assembly of coiled-coil lock washers. *Journal of the American Chemical Society* **135**, 15565–15578.

Yan Y and Erickson BW (1994) Engineering of betabellin 14D: disulfide-induced folding of a beta-sheet protein. *Protein Science* **3**, 1069–1073.

Yang J, Gustavsson AL, Haraldsson M, Karlsson G, Norberg T and Baltzer L (2017*a*) High-affinity recognition of the human C-reactive protein independent of phosphocholine. *Organic & Biomolecular Chemistry* **15**, 4644–4654.

Yang J, Koruza K, Fisher Z, Knecht W and Baltzer L (2017*b*) Improved molecular recognition of Carbonic Anhydrase IX by polypeptide conjugation to Acetazolamide. *Bioorganic & Medicinal Chemistry* **25**, 5838–5848.

Yano Y, Takemoto T, Kobayashi S, Yasui H, Sakurai H, Ohashi W, Niwa M, Futaki S, Sugiura Y and Matsuzaki K (2002) Topological stability and self-association of a completely hydrophobic model transmembrane helix in lipid bilayers. *Biochemistry* **41**, 3073–3080.

Ye S, Discher BM, Strzalka J, Xu T, Wu SP, Noy D, Kuzmenko I, Gog T, Therien MJ, Dutton PL and Blasie JK (2005) Amphiphilic four-helix bundle peptides designed for light-induced electron transfer across a soft interface. *Nano Letters* **5**, 1658–1667.

Yeates TO (2017) Geometric principles for designing highly symmetric self-assembling protein nanomaterials. *Annual Review of Biophysics* **46**, 23–42.

Yin H, Slusky JS, Berger BW, Walters RS, Vilaire G, Litvinov RI, Lear JD, Caputo GA, Bennett JS and DeGrado WF (2007) Computational design of peptides that target transmembrane helices. *Science* **315**, 1817–1822.

Yoo J, Louis JM, Gopich IV and Chung HS (2018) Three-color single-molecule FRET and fluorescence lifetime analysis of fast protein folding. *Journal of Physical Chemistry B* **122**, 11702–11720.

Yu FT, Cangelosi VM, Zastrow ML, Tegoni M, Plegaria JS, Tebo AG, Mocny CS, Ruckthong L, Qayyum H and Pecoraro VL (2014) Protein design: toward functional metalloenzymes. *Chemical Reviews* **114**, 3495–3578.

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press

Zastrow ML and Pecoraro VL (2013*a*) Designing functional metalloproteins: from structural to catalytic metal sites. *Coordination Chemistry Reviews* **257**, 2565–2588.

Zastrow ML and Pecoraro VL (2013*b*) Influence of active site location on catalytic activity in *de novo*-designed zinc metalloenzymes. *Journal of the American Chemical Society* **135**, 5895–5903.

Zastrow ML, Peacock AFA, Stuckey JA and Pecoraro VL (2012) Hydrolytic catalysis and structural stabilization in a designed metalloprotein. *Nature Chemistry* **4**, 118–123.

Zeng J, Jiang F and Wu YD (2016) Folding simulations of an alpha-helical hairpin motif alpha t alpha with residue-specific force fields. *Journal of Physical Chemistry B* **120**, 33–41.

Zhang S (2017) Discovery and design of self-assembling peptides. *Interface Focus* **7**, 20170028.

Zhang S, Holmes T, Lockshin C and Rich A (1993) Spontaneous assembly of a self-complementary oligopeptide to form a stable macroscopic membrane. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 3334–3338.

Zhang S, Yan L, Altman M, Lassle M, Nugent H, Frankel F, Lauffenburger DA, Whitesides GM and Rich A (1999) Biological surface engineering: a simple system for cell pattern formation. *Biomaterials* **20**, 1213–1220.

Zhang SQ, Kulp DW, Schramm CA, Mravic M, Samish I and DeGrado WF (2015) The membrane- and soluble-protein helix–helix interactome: similar geometry via different interactions. *Structure* **23**, 527–541.

Zhang HV, Polzer F, Haider MJ, Tian Y, Villegas JA, Kiick KL, Pochan DJ and Saven JG (2016) Computationally designed peptides for self-assembly of nanostructured lattices. *Science Advances* **2**, e1600307.

Zhang SQ, Chino M, Liu L, Tang Y, Hu X, DeGrado WF and Lombardi A (2018*a*) *De novo* design of tetranuclear transition metal clusters stabilized by hydrogen-bonded networks in helical bundles. *Journal of the American Chemical Society* **140**, 1294–1304.

Zhang SQ, Huang H, Yang J, Kratochvil HT, Lolicato M, Liu Y, Shu X, Liu L and DeGrado WF (2018*b*) Designed peptides that assemble into cross-alpha amyloid-like structures. *Nature Chemical Biology* **14**, 870–875.

Zhong QF, Jiang Q, Moore PB, Newns DM and Klein ML (1998) Molecular dynamics simulation of a synthetic ion channel. *Biophysical Journal* **74**, 3–10.

Zhou FX, Cocco MJ, Russ WP, Brunger AT and Engelman DM (2000) Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Natural Structural Biology* **7**, 154–160.

Zhou FX, Merianos HJ, Brunger AT and Engelman DM (2001) Polar residues drive association of polyleucine transmembrane helices. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 2250–2255.

Zhu Y, Alonso DO, Maki K, Huang CY, Lahr SJ, Daggett V, Roder H, DeGrado WF and Gai F (2003) Ultrafast folding of alpha3D: a *de novo* designed three-helix bundle protein. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 15486–15491.

Zozulia O, Dolan MA and Korendovych IV (2018) Catalytic peptide assemblies. *Chemical Society Reviews* **47**, 3621–3639.

https://doi.org/10.1017/S0033583519000131 Published online by Cambridge University Press