Towards functional de novo designed proteins
William M Dawson¹, Guto G Rhys¹ and Derek N Woolfson¹,²,³

Our ability to design completely de novo proteins is improving rapidly. This is true of all three main approaches to de novo protein design, which we define as: minimal, rational and computational design. Together, these have delivered a variety of protein scaffolds characterised to high resolution. This is truly impressive and a major advance from where the field was a decade or so ago. That all said, significant challenges in the field remain. Chief amongst these is the need to deliver functional de novo proteins. Such designs might include selective and/or tight binding of specified small molecules, or the catalysis of entirely new chemical transformations. We argue that, whilst progress is being made, solving such problems will require more than simply adding functional side chains to extant de novo structures. New approaches will be needed to target and build structure, stability and function simultaneously. Moreover, if we are to match the exquisite control and subtlety of natural proteins, design methods will have to incorporate multi-state modelling and dynamics. This will require more than black-box methodology, specifically increased understanding of protein conformational changes and dynamics will be needed.

Addresses
¹School of Chemistry, University of Bristol, Cantock’s Close, Bristol BS8 1TS, UK
²School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol BS8 1TD, UK
³BrisSynBio, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK

Corresponding author: Woolfson, Derek N (D.N.Woolfson@bristol.ac.uk)

Current Opinion in Chemical Biology 2019, 52:102–111
This review comes from a themed issue on Synthetic biomolecules
Edited by Johan Winne and Annemieke Madder

https://doi.org/10.1016/j.jcbpa.2019.06.011
1367-5931/© 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction
De novo protein design is said to have come of age [1]. From the early de novo proteins confirmed by high-resolution structures [2–4], the field has advanced rapidly with new scaffolds covering all-α [5,6,7], all-β [8], and mixed-α/β and α + β structural space [9,10,11]. In addition, side-chain constellations can be controlled exquisitely to introduce networks of hydrogen bonds throughout target structures [12], which, in turn, can improve the design and characterisation of de novo membrane proteins [13].

However, the ability to design functional de novo proteins from scratch, or to embellish existing de novo scaffolds with new functions, is still in its infancy. Herein, we use terms like ‘functional protein design’ for any stably folded de novo protein frameworks that incorporate interactions with small or large molecules, catalytic activity and so on. With notable exceptions—for example, reports of a functional ion transporter [14], a de novo designed catalytic triad [15*], and a highly efficient de novo enzyme [16*]— general design principles for functional protein design are sparse. Indeed, it may be that overoptimised de novo proteins, which are often hyperthermally stable, may not make good platforms for functional design, as it is known that dynamics play essential roles in ligand binding and catalysis [17–19].

Herein, we focus on truly de novo proteins rather than those achieved through protein engineering or redesign—that is, where functions are improved in or introduced to natural proteins. Of course, the latter have led to novel enzymes and ligand-binding proteins [20–22]. Whilst impressive, protein engineering relies on the inherent stability of natural scaffolds and their tolerance to mutation, and often uses the randomness of directed evolution to access the targeted function [23]. By contrast, de novo protein design removes the dependence on naturally evolved scaffolds, and has the potential for a deeper understanding of the contribution that every side chain makes towards the structure, stability and function of de novo proteins. Of course, this is an extremely challenging approach and its goals are ambitious.

Notable advances have also been made in introducing metal-binding and protein–protein interactions into de novo proteins (see recent reviews Refs. [24–26]). However, these pose different challenges to those laid out herein, and are only mentioned in passing in this review.

From minimal, through rational, to computational design
There is no single approach to protein design. However, the field can be split broadly into three different approaches (Figure 1). In minimal design binary patterns of polar (p) and hydrophobic (h) residues are used to define a target structure [27,28]. α-Helices lend themselves to this as they can be directed to fold and assemble with sequence patterns of the type hphp. As a result, the vast majority of work in this area has targeted four-helix bundles. Rational design goes a step further by incorporating more-specific sequence-to-structure
Overview of minimal, rational and computational design approaches. **Minimal design** relies on binary patterning of hydrophobic (h) and polar (p) to define a target structure. Despite considerable effort, few of these have been validated through to high-resolution structures. Nonetheless, such minimal scaffolds have been modified to introduce ligand binding and catalysis. The vast majority of minimal protein design has focused on four-helix bundle proteins. In **rational design**, which can incorporate computational methods, binary patterns are supplemented by specific sequence-to-structure relationships for the target; for example, subtly different combinations of Ile and Leu side chains in coiled-coil interfaces can direct alternate oligomer states. Such rules can be very powerful when coupled with parametric design to build, score and rank multiple models for a target. This approach has now led to many high-resolution structures including for structures not known or rare in biology; for example, a
relationships, or design rules, often garnered from inspection of the sequences and structures of natural proteins [29]. In both minimal and rational designs, extant stable scaffolds are then modified to produce functional variants. Computational design generally uses databases of structural motifs, for example, short peptide fragments, to construct the target scaffolds and to fit many primary sequences onto these [1]. In this way, large numbers of models are built and scored with an energy function. This allows variants to be ranked ahead of experimental studies. This approach also facilitates the introduction of functionality early in the design process; that is, stable proteins can be built around a target function [30**]. Another advantage of computational design over the minimal and rational approaches is that it allows access to more-complex structures [8]. That said, the combination of rational and computational approaches, particularly using parametric design to generate the backbone scaffolds, is proving powerful in delivering a variety of de novo proteins that both mimic natural protein structures and expand upon them.

The sections below build on these ideas emphasising functional designs that have been achieved thus far within each approach.

**Minimal design of functional four-helix bundles**

DeGrado, Hecht and Dutton have pioneered the concepts of minimal and rational de novo protein design (reviewed extensively in Refs. [27,33,34]). In short, these combine chemical intuition about protein structure and basic sequence-to-structure relationships to deliver straightforward designed protein scaffolds (Figure 1). Key targets in these endeavours have been four-helix bundles, which involve the coalescence of amphipathic helices encoded by self-associating peptides or within single polypeptide chains. For some time, these have been adapted to deliver functional designs.

In the ‘maquette’ approach [35], Dutton and coworkers [36,37] iteratively develop a minimal four-helix scaffold that is characterised at each step (Figure 2). Sheehan et al. use this to design a biliverdin-binding protein [37]: starting from a molten-globule state with promiscuous binding [38], potential binding sites are probed experimentally through cysteine-ligation scanning, and the resulting binding site is stabilised further by rational design.

Recently, Watkins et al. demonstrate how powerful minimalistic design can be in functional de novo design. The authors reposition heme C binding sites within a foregoing four-helix maquette (Figure 2) [39]. The resulting construct shows activity for oxidation and oxidative dehalogenation [16**]. Impressively, the kinetic analysis reveals that this de novo catalyst is as proficient as natural oxidoreductase enzymes, but with enhanced chemical and thermal stability.

Similarly, Donelly et al. apply binary patterns of polar (p) and hydrophobic (h) residues—for example, phpphpphphhp sequences—to produce a catalytic four-helix bundle from two helix-loop-helix monomers [40*]. Building on previous work to select enzyme-like functions from libraries of de novo sequences [41], the authors find one construct that hydrolyses ferric enterobactin with enantiomeric selectivity. Further investigations show that five polar/charged amino acids in the core are key to activity. This is the first example of a de novo protein that is essential for maintaining living cells, though it was achieved through selection rather than rational design.

Following a tradition established by Lear et al. [42], Lalaurie et al. employ minimal design to deliver a de novo membrane protein [43]. By analysing a small subset of natural membrane proteins, the authors develop a low-complexity leucine-rich sequence. This embeds in membranes and binds heme, although attempts at using this in catalysis appear to result in degradation of the cofactor.

Overall, the minimalistic approach to design has been successful for four-helix bundles. However, the lack of high-resolution structures for many of these designs emphasises the need to consider the stereochemical arrangement of the residues, that is, side-chain packing, to achieve well-ordered protein cores and, with these, better-defined 3D structures. For those cases where structural data have been obtained it has been for apoproteins, that is, the protein scaffold without ligand or catalytic residues/prosthetic groups present, rather than functional de novo four-helix bundles [44–47]. Arai et al. present the structure of a minimally designed four-helix bundle with primitive esterase and lipase activity [48]. However, this protein is shown to form a domain-swapped dimeric species, rather than the expected monomeric species. Further highlighting the limitations of minimal design, computational approaches have led to high-resolution structural data of both inert and functional four-helix bundles [30**,49].

(Figure 1 Legend Continued) heptameric coiled coil, CC-Hept (PDB: 4PNA). Functions have then been incorporated into these scaffolds; for example, a model of famesyl diphosphate (green) bound in a heptameric coiled coil [51]. Computational design often uses databases of protein fragments to assign thousands of potential amino-acid sequences to the design target. Energy functions are used to rank the designs with the most favourable being taken forward for experimental validation. Increasingly, functionality is being incorporated in the initial design stage rather than being appended to a stable scaffold; for example, the design of a fluorescence-activating β-barrel (PDB: 6CZI) [52**].

**Rational parametric design of functional assemblies**

Rules-based or rational protein design and computational design do not have to be mutually exclusive (Figure 1). By incorporating design rules into computational design algorithms, the number of models that need to be built and scored can be reduced dramatically. Parametric design lends itself to this. Here, target protein folds are described mathematically with a minimal number of parameters. Not surprisingly given their simplicity and potential regularity, de novo four-helix bundles have been designed parametrically [50]. However, except for a single example [30**], high-resolution structural data validating the models remain elusive.

Coiled-coil proteins also lend themselves to parameterisation. Before moving onto computational coiled-coil design, it is worth highlighting the designability of these structures because of their relatively straightforward sequences and structures. For example, Harbury et al. describe variants of the GCN4 leucine zipper with combinations of Ile and Leu residues in the core to produce parallel dimeric, trimeric and tetrameric structures, and to deliver rules for oligomer-state selection [51]; *n.b.*, wild-type GCN4 leucine zipper is a parallel homodimer. Fletcher et al. use these rules to design fully de novo homomeric dimers to tetrarmers, and Thomas et al. supplement the rules to deliver heterodimers with a range of dissociation constants [52,53]. These designs have proven useful as highly stable and robust building blocks for supramolecular assembly in materials science and synthetic biology [54–57]. However, they have no inherent function.

Crick was the first to describe coiled-coil structures parametrically [58]. Starting with the tight geometry of the α helix, he reasoned that coiled-coil structures could be defined by the radius and pitch of a superhelical assembly with two or more α helices plus a parameter (the interface angle) for the relative twist between helices (Figure 1). Numerous implementations of Crick’s equations are now available to generate coiled-coil scaffolds computationally and to build de novo sequences into these [59–65]. Thomson et al. adapt coiled-coil design principles and rules and combine them with parametric computational design to target larger discrete coiled-coil assemblies [5]. By increasing the size of the hydrophobic interface presented by the component α helices, de novo pentamers, hexamers and heptamers are achieved. These are termed α-helical barrels as they possess a central channel. Whilst these structures are not functional themselves, the fully accessible channels are prime targets for functionalisation [66]. Similarly, Huang et al. use parametric design within Rosetta to create (hyperstable) trimeric, tetrameric and pentameric coiled coils [67]. Similarly, these de novo assemblies are not functional themselves.

Burton et al. use rational design to introduce hydrolase activity into the heptameric coiled-coil scaffold, CC-Hept [15**]. In this design, each helix contributes a Cys-His-Glu catalytic triad to the lumen of the barrel (Figure 3). Kinetic analysis shows CC-Hept-CHE to be on par with other de novo and engineered catalysts, although these are all...
poor compared with natural esterases and design or engineered systems that incorporate metals [68]. This heptameric hydrolase is the first example of a functional catalytic triad incorporated into a completely de novo designed scaffold.

In addition to hydrolysing substrates, the α-helical barrels can bind other small molecules. Thomas et al. perform a systematic study to probe the size and shape of molecules that can be sequestered within the hydrophobic channels [31*]. Without modification, the pentamer, hexamer and heptamer all bind small, hydrophobic molecules with low μM affinities. Specificity for negatively or positively charged molecules has been added through the rational placement of ionisable side chains in the lumen (Figure 3).

**Fragment-based computational design beyond protein engineering**

As protein structures increase in complexity, more sophisticated approaches are needed to access more elaborate architectures. By harnessing the power of computers, thousands of designs can be generated and analysed in silico at scales beyond minimal and rational design. The most widespread approach is fragment-based design, which has three aspects: libraries of fragments or motifs are taken from structural databases, algorithms are developed to combine these to assemble target structures, and scoring functions are used to assess both the assembled structures and sequences that best fit onto them (Figure 1) [69–72]. This is epitomised by the Rosetta suite for computational protein design developed by the Baker group [73].

There are numerous examples of new functions being engineered into natural proteins using these methods [74–76], including opioid binders [77], an amino-acid binder [78] and Schiff-base-forming enzyme [79]. A related approach mimics nature by combining larger protein fragments [80] and has proven successful for generating non-functional de novo proteins [81,82]. Whist relying heavily on the evolutionary traits of the parent enzymes, these chimeric proteins have activities that match their natural counterparts. Lapidoth et al. adapt this approach in an automated fashion to create TIM barrels, a ubiquitous fold consisting of eight α-helices and eight β-strands arranged in tandem, that is, (βα)₈, with hydrolase and lactonase activity [83*].

Huang et al. and Marcos et al. have designed de novo proteins incorporating cavities with potential for catalysis or small-molecule binding [10*,11]. In the first study, a de novo four-fold symmetric (βα)₄-barrel is designed using RosettaRemodel developed for repeat proteins [10*]. This is of interest as TIM barrels are the most common enzyme topology found in nature. The second study develops design principles for curved β sheets [11]. Applying analyses of bulges and register shifts in naturally curved β sheets, the authors use RosettaDesign to obtain nine de novo scaffolds with pockets that could be modified for ligand binding. Serendipitously, the crystal structure of one scaffold has a ligand bound in the cavity, highlighting the potential for functionalisation.

Despite these successes, fragment-based design might be considered a ‘black-box approach’ with few design rules or general principles being gleaned. For example,
impressively, Rocklin et al. apply a massive-scale approach to protein design, coupling stability against protease degradation with yeast display to deliver a large number of stable, de novo mini-proteins [84]. That said, a design rule to emerge from this study is that certain charged side chains near the termini of helices stabilise the constructs, which is in agreement with conclusions drawn from a previous study combining bioinformatics and rational design of single α helices [85].

**Designing in function from the beginning**

Most of the studies described above focus on the design of the de novo scaffold before functionalisation or improving an already established functional de novo protein. Though there are clear examples to the contrary [15**,31*,39], extant de novo designed scaffolds may not have suitable cavities or sites for every targeted function, or the post-incorporation of the necessary functional residues may prove problematic. Indeed, for natural proteins it is well documented that small changes, even distal to the binding site/active site, can have large effects on the function [86]. Therefore, design strategies that incorporate, or at least consider, the functional aspect at an early stage could ultimately lead to more successful outcomes.

Polizzi et al. describe such an approach to design a porphyrin-binding four-helix bundle. Of course, four-helix bundles that bind porphyrins have been designed previously. In fact, tight binding to a porphyrin cofactor in de novo four-helix bundles is common due to the hydrophobicity of the ligand and the strengths of side chain-metal interactions [87,88]. However, the lack of structural data from these studies has precluded validation of these designs. With this in mind, Polizzi et al. simultaneously design a well-folded hydrophobic core and a ligand-binding site into a four-helix bundle (Figure 4) [30**]. By factoring the long-range influence of residues distal to the ligand binding site, the authors improve on earlier designs [89] and obtain a high-resolution structure.

Dou et al. take a similar approach to design ligand-binding β-barrel proteins [32**]. Recognising irregularities in sheets, the authors use a 2D map of side-chain interactions and ‘kinks’ in the structure caused by glycine residues to direct 3D model building. This results in the successful design, characterisation and crystallisation of the first de novo water-soluble β-barrel protein. Furthermore, β barrels that bind small molecules are targeted to incorporate an environment-sensitive fluorophore that only fluoresces when held in a specific conformation by the de novo scaffold (Figure 4). The fluorophore is introduced early in the design strategy, rather than by embellishing a non-functional variant. This study is impressive for two reasons: firstly, accessing soluble β-rich proteins has proven challenging in protein design; secondly, the de novo proteins activate fluorescence of the small molecules in vivo. That said, before library screening is used to improve the designs, the low μM affinity of the small molecule is similar to previously reported binding constants to α-helical barrels [31*].

**Challenges ahead**

The robust and routine design of functional de novo proteins remains an unsolved problem. For instance, to our knowledge, there are no examples of tight binding of small, polar molecules by de novo proteins. The change in approach in the last few years to incorporate the functional aspect of the design at an early stage shows clear potential, which we envisage will be become more evident as design algorithms improve. However, as stated above, accessing functional de novo proteins that work on a par with natural proteins will likely require the incorporation of conformational changes and dynamics into the design process [90]. Such design targets will need improved abilities to build and

---

**Figure 4**

Computationally designed functional proteins. (a) The first high-resolution structure of a porphyrin-binding de novo four-helix bundle [30**]. The interaction between the zinc atom in the unnatural porphyrin ring (C₆H₄F₂N₅O₂Zn) and the histidine side chain is shown (PDB: 5TGy). (b) An example of a fully de novo water soluble β-barrel (PDB: 6CZH) [32**]. The environmental sensitive ligand, DFHBI (C₃H₅F₃N₂O₂) bound to the cavity (green spheres).
score in silico models that access multiple states. Thankfully, methods for multistate design are being developed [91–94]. For instance, Grigoryan et al. use such an approach to design leucine zippers that selectively bind a single partner from 20 members of the bZIP family by modelling potential off-target interactions as part of the design process [95]; Löfler et al. engineer a (βα)₆-barrel into a retro-aldolase with measurable, albeit low, catalytic efficiency [94]; and Feng et al. use conformation ensembles to engineer ligand-binding G-protein-coupled receptors [96].

Natural allosteric proteins can be engineered to bind different small molecules [76,97]. Similarly, existing allosteric systems can be used to control new functions [98,99]. Switchable de novo coiled-coil systems, both reversible and irreversible, can be controlled through temperature [100], pH [101,102] and metal binding [103–105]. However, fully de novo allosteric proteins that respond to small-molecule inducers have yet to be reported.

Arguably more progress is being made utilising dynamic multistate de novo design. Davey et al. recently give an example of an, albeit engineered, dynamic protein that accesses two conformations that exchange on a millisecond timescale [106]. Focusing on de novo proteins, Rhys et al. present a de novo α-helical barrel that is hexameric in solution but crystallises as an octameric assembly [6]. Joth et al. present a de novo zinc-ion transporter by designing a membrane-spanning four-helix bundle with two distinct coordination sites each of which destabilises the other upon metal binding [14]. However, the challenge of incorporating dynamics to improve catalysis or small-molecule binding has yet to be met.

Overall, despite considerable and encouraging advances in de novo protein design there are many challenges ahead for the de novo design of functional proteins. These are being actively targeted by the field as a whole. If advances continue at the current rate of delivery of de novo protein scaffolds, then protein design will indeed have come of age.

Conflict of interest statement
Nothing declared.

Acknowledgements
WMID, GGR and DNW were supported by a European Research Council Advanced Grant (346764). DNW is supported by the EPSRC and BBSRC through the BmSynBio Synthetic Biology Research Centre (BB/L01386X1). DNW holds a Royal Society Wolfson Research Merit Award (WM140008). We thank Ross Anderson for providing a computational model of the C4S protein.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest


This paper describes an empirical investigation of coiled coils with greater than five helices. A stable, multi-state assembly is presented, and also the first structure of an octameric coiled coil.

11. The first example of a structurally validated computationally designed TIM barrel. Although not functional, this is the most common topology for natural enzymes and shows the potential for the scaffold to be functionally established in the future.
31. Computational design is used to design stability and binding for an unnatural porphyrin in a four-helix bundle simultaneously. This is the first example of a high-resolution structure for this class of de novo protein with its ligand bound.
33. A systematic study of ligand binding to the channels of de novo pentameric, hexameric and heptameric α-helical barrels. Ionizable amino acids are introduced to the lumens of the barrels to enable binding of positively and negatively charged molecules.
35. De novo designed water-soluble β barrels are described for the first time. Variants that bind an environmentally sensitive fluorophore are characterised both in vitro and in vivo.
44. A minimal protein design is embalished using combinatorial libraries to produce a four-helix bundle capable of mimicking the function of an essential enzyme.

Functional de novo proteins Dawson, Rhys and Woolfson 109
110 Synthetic biomolecules


Computational design is used to combine fragments from enzyme families and install xylanase and lactonase activity. Activities matching the natural proteins are obtained without the need to evolve the sequences further.


A combination of protein design, homology modelling and ligand docking is used to predict GPCR-ligand conformations and to engineer new ligand-binding specificity into a receptor of unknown structure.


A natural protein is engineered to switch between two distinct conformations on a millisecond timescale. This is the first example of dynamics being incorporated by rational protein redesign.