Review Article

A Brief History of *De Novo* Protein Design: Minimal, Rational, and Computational

Derek N. Woolfson *

School of Chemistry, *University of Bristol, Cantock's Close, Bristol BS8 1TS, UK*
School of Biochemistry, *University of Bristol, Biomedical Sciences Building, University Walk, Bristol BS8 1TD, UK*
Bristol BioDesign Institute, *University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK*

Correspondence to Derek N. Woolfson: *d.n.woolfson@bristol.ac.uk*
https://doi.org/10.1016/j.jmb.2021.167160
Edited by Louise C. Serpell

Abstract

Protein design has come of age, but how will it mature? In the 1980s and the 1990s, the primary motivation for *de novo* protein design was to test our understanding of the informational aspect of the protein-folding problem; *i.e.*, how does protein sequence determine protein structure and function? This necessitated minimal and rational design approaches whereby the placement of each residue in a design was reasoned using chemical principles and/or biochemical knowledge. At that time, though with some notable exceptions, the use of computers to aid design was not widespread. Over the past two decades, the tables have turned and computational protein design is firmly established. Here, I illustrate this progress through a timeline of *de novo* protein structures that have been solved to atomic resolution and deposited in the Protein Data Bank. From this, it is clear that the impact of rational and computational design has been considerable: More-complex and more-sophisticated designs are being targeted with many being resolved to atomic resolution. Furthermore, our ability to generate and manipulate synthetic proteins has advanced to a point where they are providing realistic alternatives to natural protein functions for applications both *in vitro* and in cells. Also, and increasingly, computational protein design is becoming accessible to non-specialists. This all begs the questions: Is there still a place for minimal and rational design approaches? And, what challenges lie ahead for the burgeoning field of *de novo* protein design as a whole?

Introduction from protein engineering to protein design

In August 1988 I attended the Protein Society's 2nd Annual Symposium in San Diego.[1] I was a first-year graduate student, it was my first international conference, and it changed my life. Bill DeGrado spoke about his attempts to design four-helix bundle proteins using a chemocentric approach to assemble short amphipathic helices in water.[2,3] I was hooked, and I wanted to design proteins from then on. The intervening 30+ years have witnessed considerable changes in both the approaches to and the successes of this field of *de novo* protein design. Indeed, the early minimal and rational approaches to protein design—which were pioneered by a small number of research groups—have been overtaken by computational protein design, which is now being practiced by

Please cite tts article as: D.N. Woolfson, A Brief History of *De Novo* Protein Design: Minimal, Rational, and ComputationalJournal of Molecular Biology, https://doi.org/10.1016/j.jmb.2021.167160

many more researchers. This leads to questions that I pose here: have minimal and rational *de novo* protein design served their purpose? Is it time to move on and fully embrace the power and potential of computational *de novo* protein design? And, what challenges remain for the field? Before I address these questions, I'd like to make it clear that this article is not intended as an exhaustive review of the *de novo* protein design field – that would be challenging. Rather, it is a more-focussed piece written from my own perspective. There are several exceptional recent reviews of protein design, which together give a thorough overview of the current state-of-the-art of the field.[4–7] Also, I'd like to set the scene in three ways as follows.

First, I need to make a distinction. Protein design is different from protein engineering. In the latter, natural proteins are engineered or redesigned to understand natural protein functions, to alter these functions, or to repurpose the natural protein structures for new functions. The advantage is that protein structure, stability and dynamics are defined, even if they are not fully understood. The basic tenet of protein engineering is that these features remain largely unperturbed following mutagenesis. This allows researchers to target altered or new functions through rational mutagenesis or by directed evolution.[8,9] The impact of protein engineering over the past four decades has been astounding, and this was acknowledged by the award of the 2018 Nobel Prize for Chemistry to Frances Arnold, George Smith and Greg Winter.[10]

That said, and before moving on from protein engineering, I should note that considerable progress in this field has been made by protein designers developing computational tools to guide mutagenesis studies. David Baker's Rosetta suite of computational tools being a prime example[11,12]; and an excellent specific example is Sarel Fleishman's PROSS tool and server for improving protein expression and stability through computationally selected mutations.[13] In addition, improved computational methods have contributed to advances in enzyme design. At present, though with some exceptions,[14,15] these have largely been applied to repurposing natural protein structures. I do not review this exciting area here, but others who have pioneered the field—notably, Don Hilvert and Ross Anderson—have done so.[8,16,17] Thus, in these respects the boundary between protein engineering and protein design is blurred. Indeed, I recognise fully that making wholesale changes to natural proteins to introduce completely new functions—such as catalysis, small-molecule binding, and protein–protein interactions—is non-trivial and is considered *protein design* by many.

By contrast, in *de novo* protein design it is generally accepted that the starting point or scaffold isn't a given. Indeed, it *is* the initial target for design. Usually, and in the broadest terms, this is done by choosing a target protein shape or assembly and then finding a sequence, or sequences most likely to fold to and stabilise that target. For the next step of introducing function, there are two schools of thought in *de novo* design[18]: some advocate generating stably folded scaffolds first and then embellishing these with functions; while others argue that the processes of structural and functional design should be more integrated–in short, the *de novo* proteins should be built around the functional regions.

Second, and before getting too embroiled in the details of protein design, we need some boundaries, definitions and illustrations. The primary objective and indeed the definition of *de novo* protein design is simple: *what synthetic sequence(s) can be made to direct the folding and stabilisation of a target protein structure and function?* As such, protein design is often called the inverse protein-folding problem.[19] Of course, the reality is that this is far from simple to achieve; indeed, it has proved extremely challenging to get to where we are today. In Box 1, I give definitions for what I consider the three main approaches to protein design—namely, *minimal, rational*, which encompasses *consensus*, and *computational design*—together with some of their pioneers, and the merits and shortcomings of the different approaches. I find the definitions useful, and I hope that they would be generally accepted by the broader protein-design community. In Figures 1 and 2, which I describe in detail below and refer to throughout the article, I have attempted to plot a *structural history* of the progress made in *de novo* protein design using these approaches. This is a timeline of experimentally determined high-resolution structures of *de novo* peptides and proteins. My intention is to illustrate the trajectory and progress of the field over three decades. My hope is that it will form the basis for building a structural database of *de novo* protein designs. In turn, this would be a source for viewing, analysing and learning from *de novo* protein designs, as has been achieved using the RCSB PDB and relational databases that have sprung from it over five decades.[20]

Box 1. Approaches and pioneers.

Minimal protein design[6,21] uses straightforward chemical principles such as patterning of polar (*p*) and hydrophobic (*h*) amino-acid residues to direct the folding and assembly of secondary structures leveraging the hydrophobic effect. For instance, in aqueous buffers ...*hpphppp*... and ...*hphphp*... patterns can lead to the formation and association of amphipathic a helices and b strands, respectively. The experimental approaches to this can be step-wise—*i.e.,* through careful biophysical studies of small numbers of synthetic polypeptides, as pioneered by DeGrado.[7] Or they can be combinatorial—*i.e. via* the generation of libraries of large numbers of protein sequences from synthetic genes built with redundant codons, and then selecting variants with desired biophysical or functional properties, as championed by Michael Hecht.[22,23] The advantage of these approaches is that they test straightforward concepts and principles, and that complexity can be built from the bottom up. Other early proponents of minimal design include: Jane and David Richardson[24–26]; Les Dutton who pioneered the maquette approach to design functional helical bundles that incorporate various cofactors[27,28]; and Bob Hodges who designed the first *de novo* coiled-coil peptides.[29,30] Sadly, Bob passed away recently.[31] The minimal approach resulted in the concepts of positive and negative protein design: *i.e.*, that even seemingly straightforward design targets such as the 4-helix bundle can have multiple arrangements; therefore, whilst positive design principles are needed to design towards the target, negative design is also required to direct away from unwanted alternative states.

Though the boundary is somewhat blurred, **rational protein design**[3,32] builds on from minimal approaches by embellishing simple *hp* patterns with more-specific sequence-to-structure relationships garnered from biochemical, bioinformatics or empirical studies. The advantage of speed of turnaround in design-build-test cycles from minimal design is maintained, but the robustness of the designs is improved by including biochemical and evolutionary sequence-to-structure data. One of the first examples to bridge minimal and rational protein design was Lynne Regan and DeGrado's single-chain 4-helix bundle,[33] which Regan and Neil Clarke adapted to introduce a zinc-binding site.[34] The rational approach has worked particularly well for targets such as multi-a-helix coiled-coil assemblies, where clear sequence-to-structure relationships to define different coiled-coil architectures came through the analysis of natural structures and systematic experiments.[35–37] Early exponents here include: Tom Alber,[38,39] Vince Conticello,[40,41] DeGrado,[7,42] Hodges,[30,43] Peter Kim,[44] Vince Pecoraro,[45] and my own group[46–48]; and for early b structures Tanja Kortemme and Luis Serrano.[49]

I also include **consensus protein design** in the rational design category.[50–52] In this approach, multiple alignment of natural sequences for a target structure are used to identify the key residues and residue spacings that define the fold. These are then used to build synthetic versions of the structure. A difference between the consensus and the general rational approaches is that the former can only mimic the natural target structures, whereas the latter can potentially be used to expand into new structures, *i.e.,* the dark matter of protein-structure space.[53,54]

There are potential downsides from evolutionary bias in the consensus approach[55]; indeed, it may even be hard to distinguish designed from natural sequences. Nonetheless, it has worked well for the design of miniproteins and repeat proteins, which can then be used as scaffolds for displaying residues that confer function such as binding.[4,56] Pioneers of this approach include: Jeremy Berg[57,58]; Barbara Imperiali,[59] Zheng-yu Peng[51,60]; Andreas Plückthun,[56,61] and Regan,[4,62] with notable contributions from Ashley Buckle,[63] Martin Lehmann,[64] Philippe Minard,[65,66] and Oliver Rackham.[67]

Computational protein design[5,7,68] generates and evaluates full atomistic models for many different sequences for a given design target *in silico* ahead of experimental studies. The backbones can be fixed or flexible, and can be generated by direct user input, fragment-based approaches or parametrically. Side chains are then added using algorithms that explore and test different rotameric states. The resulting atomistic models are scored using relatively simplified forcefields or heuristics that relate sequence, structure and internal energy. Therefore, the success of these approaches depends on generating good starting models for the backbone, rapid but reliable methods for adding side chains, and the ability of the forcefields and assessment methods to capture the salient features of protein stability and function. However, none of these are foolproof. Therefore, and increasingly, computational design is used to guide the design of relatively small sequence libraries of genes (usually hundreds to thousands), which can be made by DNA synthesis, and then expressed and screened or selected from to deliver stable or functional designs. Early pioneers of computational protein design include: Alber[38] with Pehr Harbury and Kim,[69] DeGrado[70]; Tracey Handel,[71] and Steve Mayo.[68,72] The recent rise of computational design has largely been due to Baker's group and the development of Rosetta.[5,12,73] The approach is being adopted widely by Baker, his former co-workers and others alike.[74,75] Other computational protein design platforms include OSPREY from Bruce Donald,[76,77] and ISAMBARD from Chris Wood, Drew Thomson and my own lab.[78]

N.B. In their excellent retrospective of *de novo* protein design, Ivan Korendovych and DeGrado give a more-detailed and fascinating review of the field from its inception to current-day successes, challenges and promise.[7] They plot a helpful timeline of the development of the field and its key milestones. This is different but complementary to the structure-based timeline that I present here. They also define three eras of protein design that broadly overlap with my own definitions: what I call *minimal protein design*, they refer to as *manual protein design*; where I use *rational protein design*, they use *computational design guided by fundamental physicochemical principles*, which is perhaps where our definitions differ the most; and what I call *computational protein design*, they expand to *fragment-based and bioinformatically informed computational protein design*.

Third, we should ask ourselves a tough question: *given that nature offers a wealth of structures and*

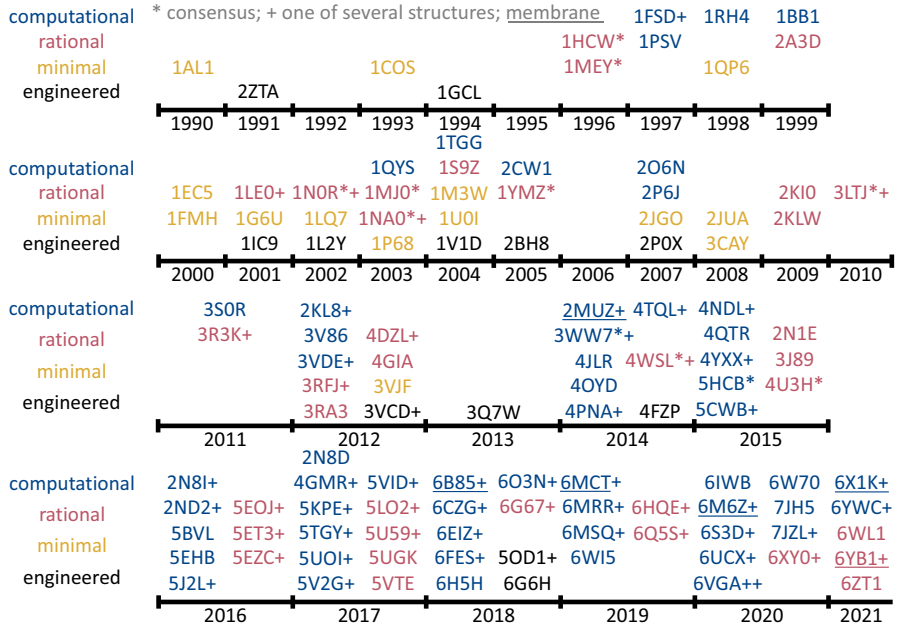


Figure 1. A structural timeline for *de novo* protein design. The timeline is based on the year of publication of primary papers that cite high-resolution *de novo* designed peptide and protein structures determined by X-ray crystallography, NMR spectroscopy or, in a small number of cases, by cryo-electron microscopy (*e.g.* 6HQE,[83] 6M6Z,[84] and 6WL1[85]). The structures are represented by their 4-letter codes from the RCSB PBD (https://www.rcsb.org/).[20]. The design categories—minimal (yellow), rational (red), consensus (*), and computational (blue)—are defined in Box 1. In addition, in black, a number of structures of "engineered" peptides and proteins have been included, as they have had an impact on *de novo* design, or are landmarks in the space between protein engineering and design. These are for the GCN4 leucine zipper (2ZTA[86]); trimeric and tetrameric variants of the leucine zipper (1GCL[87]); an engineered β-sheet miniprotein (1IC9[88]) the Trp-Cage (1L2Y[89]); a trimmed-down miniprotein with esterase activity (1V1D[90]); a selected combinatorial protein (2BH8[91]); a selected ATP binding protein (2P0X[92]); a cage complex computationally designed from natural proteins (3VCD[93]); a protein engineered using a reduced amino-acid alphabet (3Q7W[94]); a computationally engineered uranyl-binding protein (4FZP[95]); a zinc-binding assembly evolved to rapidly catalyse ester hydrolysis (5OD1[15]) from a computationally designed *de novo* metal-binding dimer MID1 (3VDE[96]); and an engineered variant of a foregoing computational design with a different topology (6G6H[97]). Key: *, consensus design; +, one of multiple, related structures from the same publication; and underlined, membrane-spanning designs. The list was compiled by manual inspection of the PDB using a combination of keyword and author searches, and knowledge of the field. As such, it is somewhat subjective and may not be complete.

functions and our success in engineering these, why design synthetic proteins at all? In the early minimalist days of *de novo* protein design, the aim was simply to test and develop our understanding of sequence-to-structure relationships in proteins by building albeit simple proteins from scratch. In other words, it was a test of our understanding of what Tom Creighton called the informational aspect of the protein-folding problem.[79] This fitted Richard Feynman's challenge, "*What I cannot create, I do not understand*". With some notable exceptions,[80–82] the idea of making synthetic proteins that might be functional or useful was far from designers' minds. I am sure that this contributed to scepticism of the emerging field by other protein scientists, which protein designers perceived in the 1980s and 1990s.[7] Indeed, early on, the approaches were largely minimal with relatively straightforward concepts being tested; for example, the pioneering work from the DeGrado and the Hecht groups established the role of patterning of hydrophobic and polar residues (*hp* patterns) to define secondary, tertiary and quaternary structures.[3,23] Minimal design is still pursued with some success and it has its place,[6] but the design landscape has changed radically. The field moved to rational and computational design approaches, which I will focus on for the remainder of this perspective. To bring this full circle, I believe that advances here are now putting us in a strong position to answer that tricky question of *why design synthetic proteins at all?* However, for *de novo* protein design to be established as a parallel to protein engineering, let alone as an alternative to it, hurdles remain in the continued development of the field.

A brief history of protein design through structures

Before I move onto the main body of this perspective, I will describe its two main figures. These come with a caution: They are a structural

viewpoint of protein design over the past three decades. Thus, they are biased to designs that have been amenable to high-resolution X-ray protein crystallography, NMR spectroscopy or, more recently, cryo-electron microscopy; and they overlook a large number of excellent designs that are thoroughly characterised in solution, but which have not been resolved to atomic resolution. Andreas Plückthun once put it to me *that the field is obsessed with RMSD and we need to get on with designing for function, if we want to be relevant*. By this he meant that the community has been concerned mostly with the RMSD (root mean squared deviation) between computational models and experimental structures, and that there are other tests of design success. For instance, how good is a design's function—*e.g.*, for binding or catalysis—compared with natural proteins? Notwithstanding this view, I find the structural history a useful framework to discuss the evolution of the field.

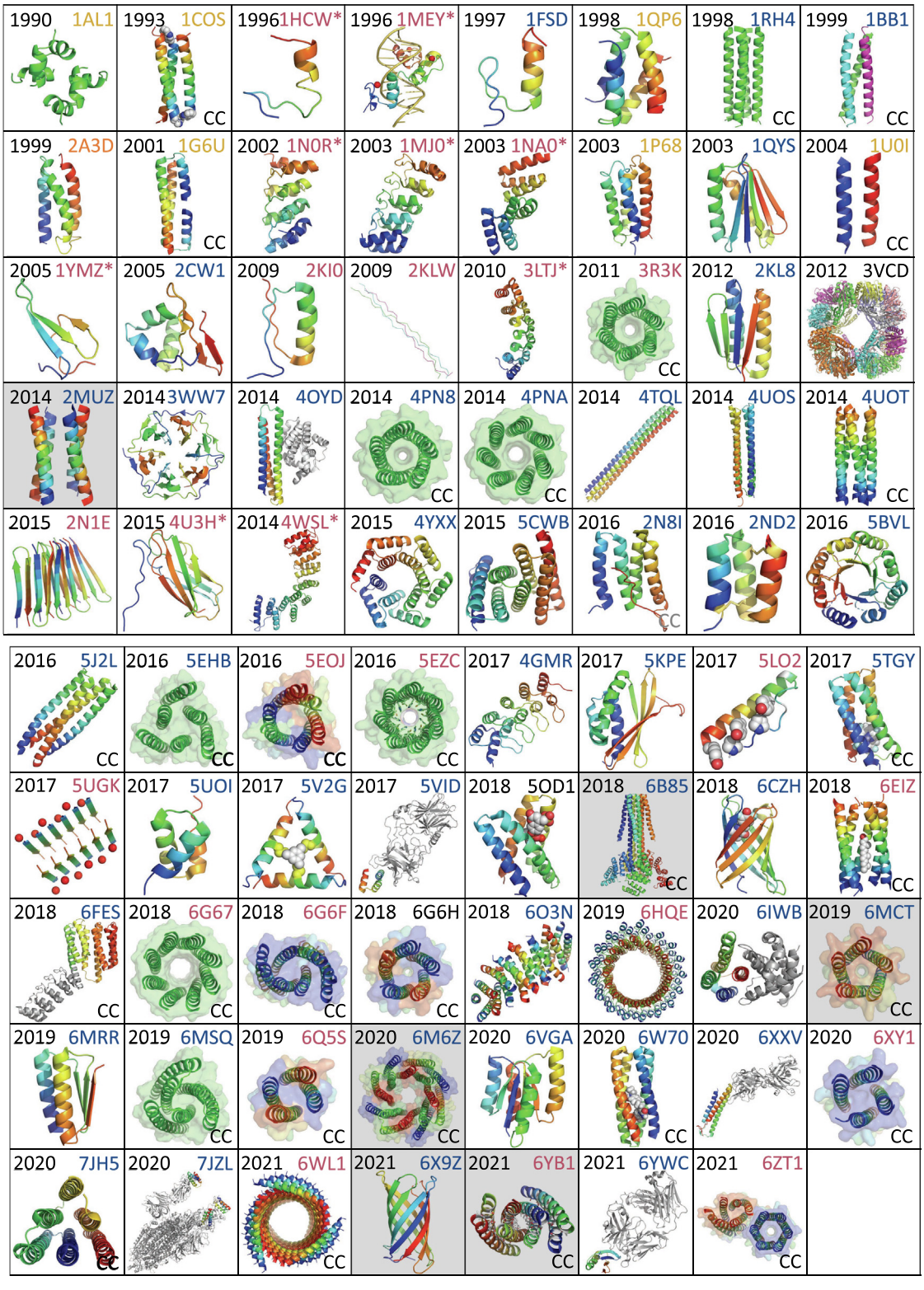
N.B. Baker and colleagues have produced their own snapshot of structurally resolved computational protein designs, which I refer readers to.[5]

Figure 1 gives a timeline with the 4-letter RCSB PDB codes for *de novo* protein designs that have been determined to atomic resolution by X-ray protein crystallography, NMR spectroscopy, or cryo-electron microscopy.[20] To illustrate points that I make below, I call out these 4-letter codes frequently in the text. Readers can view some of these in Figure 2 and the others by typing the codes into the search field of the URL, https://www.rcsb.org/. This gallery of structures was culled by hand through a series of author-name plus keyword searches and knowledge of the field, and then sifted by manual inspection. This was necessary because, at present, there is no reliable automated way to search for *de novo* protein designs in the PDB. Therefore, Figure 1 is somewhat subjective and most likely incomplete. Nonetheless, there are several take-home messages from it.

First, the number high-resolution structures of *de novo* designed peptides and proteins is increasing exponentially: roughly, this doubled from the 1990s to the 2000s, and tripled from the 2000s to the 2010s. For comparison, the whole PDB is doubling in size approximately every 6–7 years. There are now over 100 structures of *de novo* peptides and proteins, which is a good resource for the aforementioned bioinformatic analyses. Second, the plot shows a transition in approaches taken to protein design: The early minimal and engineering approaches (coloured dark yellow and black) have given way to rational design (red) and increasing to computational design (blue). Third, grouped within rational design, there was a particularly successful push in **consensus design**—defined in Box 1, and marked with *s in Figures 1 and 2—during the 2000s led by Plückthun, Regan and others[52,56]. Fourth, in terms of structures determined over the last decade, rational and computational design appear to be on a par; though the expectation is that the balance will continue to shift towards computational design. Finally, in both approaches, studies are increasingly being supported by multiple high-resolution structures (marked '+' in Figure 1), which probably reflects both the increased accessibility of structure determination and a raising of the bar in *de novo* protein design.

Figure 2 shows ribbon diagrams for representative examples of these structures. Although I have attempted to cover as much of the protein-design landscape as possible, this is not meant to be a definitive collection of structures achieved to date. I do hope that will follow. The first thing to notice is that the early designs are relatively short polypeptides with simple structures. A rough metric to guesstimate the length of a designed sequence from these cartoons is that a turn of α helix spans ≈3.5 residues. Thus, with the exception of alpha3D (2A3D[98]), the structures from the 1990s are for peptides of $\leq$ 30 amino acids. This was for understandable reasons: most of these designs were generated through minimal, rational, or very early computational design, and realised experimentally *via* solid-phase peptide synthesis. In the 2000s, the growing ambitions of designers, increasing computer power, accessibility of recombinant DNA technology, and falling costs of synthetic DNA, all contributed to the increased complexity of design targets shown by the cartoons. This is particularly apparent from 2015 onwards where the lengths and complexities of structurally resolved *de novo* proteins increasingly mirror those of natural proteins.

A second observation from Figure 2 is that—like JBS Haldane's quip that "*[God] has a special preference for beetles*"—protein designers have a fondness for α helices. This makes sense too: The α helix was an established model for protein-folding studies by Buzz Baldwin and others in the 1980s and 1990s.[112] This led to good sequence-to-structure relationships and chemical principles for building with this secondary structure. Moreover, it is 'self-contained' with backbone hydrogen-bonding potential satisfied internally, unlike β strands that must do this by assembling and often aggregating into β sheets.[113,114] Thus, with the α helix, designers can focus on using side-chain interactions to drive the folding of tertiary structures and/or the assembly of quaternary structures. This emphasis on helical assemblies is changing with advanced computational designs from the Baker group for instance, leading to structures of both water-soluble (6CZH[115]) and membrane-spanning (6X9Z[109]) β-structured proteins. Further on α-helical assemblies, however, it is interesting to compare those achieved through rational and com-


1990 1AL1
1993 1COS CC
1996 1HCW*
1996 1MEY*
1997 1FSD
1998 1QP6
1998 1RH4 CC
1999 1BB1 CC
1999 2A3D
2001 1G6U CC
2002 1N0R*
2003 1MJ0*
2003 1NA0*
2003 1P68
2003 1QYS
2004 1U0I CC
2005 1YMZ*
2005 2CW1
2009 2KI0
2009 2KLW
2010 3LTJ*
2011 3R3K CC
2012 2KL8
2012 3VCD
2014 2MUZ
2014 3WW7
2014 4OYD
2014 4PN8 CC
2014 4PNA CC
2014 4TQL
2014 4UOS
2014 4UOT CC
2015 2N1E
2015 4U3H*
2014 4WSL*
2015 4YXX
2015 5CWB
2016 2N8I CC
2016 2ND2
2016 5BVL
2016 5J2L CC
2016 5EHB CC
2016 5EOJ CC
2016 5EZC CC
2017 4GMR
2017 5KPE
2017 5LO2
2017 5TGY CC
2017 5UGK
2017 5UOI
2017 5V2G
2017 5VID
2018 5OD1
2018 6B85 CC
2018 6CZH
2018 6EIZ CC
2018 6FES CC
2018 6G67 CC
2018 6G6F CC
2018 6G6H CC
2018 6O3N
2019 6HQE
2020 6IWB
2019 6MCT CC
2019 6MRR
2019 6MSQ CC
2019 6Q5S CC
2020 6M6Z CC
2020 6VGA
2020 6W70 CC
2020 6XXV
2020 6XY1 CC
2020 7JH5 CC
2020 7JZL
2021 6WL1 CC
2021 6X9Z
2021 6YB1 CC
2021 6YWC
2021 6ZT1 CC

putational design. The former tend to have short and curved α helices, which reflect those found in natural globular and fibrillar proteins; whereas, some of the computational designs use longer and straighter helices (see 4TQL[116] and 6B85[107] for example). Finally, as with natural proteins, there are many fewer structures determined of *de novo* membrane-spanning proteins (shaded boxes in Figure 2) than there are for water-soluble proteins. Thus, the perennial challenge of working with natural transmembrane proteins extends into the design space.[6,80,84,107,108,110]

Evolution in protein design: From minimal to rational approaches

As stated above, *de novo* protein design necessarily began with minimal approaches.[2,3,7,23] Mostly, these involved encoding synthetic polypeptides with patterns of hydrophobic (*h*) and polar (*p*) residues spaced to match the helical periodicities of the two main secondary structure types: *i.e.*, variants of *hpphppp* with hydrophobic side chains placed 3 and 4 residues apart to best match the 3.6 residues per turn of the α helix; and *hphp* patterns with hydrophobic side chains every other residue to match the 2 residues per turn of the β strand. Often, this also employed reduced amino-acid palettes, with leucine as the favoured hydrophobic residue, and lysine and glutamic acid the favoured polar residues. The aim was for these sequences to fold into amphipathic secondary structures that would assemble in water driven by the hydrophobic effect to form tertiary and quaternary structures.

Despite this simplicity, some of these studies led to resolved protein structures, labelled yellow in Figures 1 and 2. Some of these structures were different from what was originally intended. For example, DeGrado's alpha1 peptide was conceived to form a 4-helix bundle,[3] but the X-ray crystal structure indicated that a hexamer was also accessible (1AL1[117]); and Hodges' early α-helical coiled-coil design targeting a parallel dimer[29] was adapted by DeGrado and Karyn O'Neill[118] and later shown to form a mixed parallel/antiparallel trimer (1COS[42]). There is a long-standing and important debate over the definitions and degrees of success and failure in *de novo* protein design,[119] and, as the adages go, we may well learn more from the latter. Indeed, these early approaches taught us some important lessons. Not least, that simple chemico-physical principles can be abstracted from natural proteins and used as rules to deliver albeit rudimentary protein-like structures. For me, the most important lesson to emerge from this approach or era[7] was the concept of *negative protein design*. In positive protein design, principles and rules are used to design *towards* a target structure. But the shortcomings of some of these early designs made the community realise that there was also the need to design *away from* alternate states that lie close in energy to the target state. As we will see, this concept is as important today as it was 30 years ago.

The change towards modern protein design occurred in the 1990s when people began to use evolutionary and empirically derived information to relate protein sequences and structures better and beyond simple *hp* patterns. I see this as a natural transition rather than as a step-change in the field. Before that, the complexity of protein sequences and structures, and the implications that this had for solving the protein-folding problem were all too apparent. For example, in protein-structure prediction the great hope was that computational approaches such as homology modelling, threading, or fragment-based structure prediction would crack that problem. That is, rather than a series of rules or heuristics for folding protein chains, a more-global view employing advanced computational algorithms would be needed. This has evolved from using amino-acid profiles or position specific scoring matrices,[120–122] through fragment-based assembly,[123] and onto advanced-statistical or machine-learning methods.[124] In turn, this has led to good pipelines for homology modelling,[125] and the AI-based protein-fold predictions epitomised by AlphaFold[124] and its success in CASP (Critical Assessment of methods for protein Structure Prediction).[126,127] Returning to protein design, better understanding of sequence-to-structure relationships was essential to deliver better rules for protein design and to advance the field. This spawned rational *de novo* protein design in the 1990s.

Relatively straightforward miniproteins like zinc fingers and leucine zippers provided inspiration in

Figure 2. A gallery of high-resolution *de novo* designed peptide and protein structures. The gallery shows selected and representative structures from the structural timeline of Figure 1. The dates and colouring scheme for the 4-letter codes carry over from Figure 1. Structures are represented as backbone cartoons rendered in PyMol (https://pymol.org/2/). Mostly, these are coloured "chainbow", *i.e.* blue to red from the *N* terminus to the *C* terminus. Some images—mostly, for parallel homomeric helical assemblies—are coloured solid green for clarity. Where additional protein chains are shown in grey these are for protein fusions to the designs (6FES[99]) or with targeted protein-protein interactions (4OYD,[100] 5VID,[101] 6IWB,[102] 6XXV,[103] 7JZL,[104] 6YWC[105]). Grey-shaded boxes indicate transmembrane structures (2MUZ,[106] 6B85,[107] 6MCT,[108] 6M6Z,[84] 6X9Z,[109] and 6YB1[110]). CC in the bottom right-hand corner of a box indicates that the *de novo* designed structure tests positive for knobs-into-holes packing characteristic of coiled-coil structures using the program SOCKET.[111]

the quest for clear and robust sequence-to-structure relationships and early design targets. In natural examples, small numbers of key residues act as keystones that largely determine the target protein structures. For example, the critical residues in one class of zinc fingers include 4-residue combinations of Cys and His residues that bind the zinc ion. Metal binding dictates the overall fold of the protein, which is stabilised further by the formation of a small hydrophobic core centred on key hydrophobic and usually aromatic residues.[128,129] Moving into design, using a consensus approach Berg and co-workers designed DNA-binding zinc-finger proteins (1MEY[58]).[57,130] There were also early and successful attempts to replace the zinc with an expanded hydrophobic core using both consensus and computational design, notably from Imperiali (1HCW*[59]) and Mayo (1FSD[72]), respectively.

More broadly, and seeded by early successes such as the Trp zipper (1LE0[131]) and the Trp cage (1L2Y[89]), various stably folded *de novo* miniproteins have been successfully targeted: through rational design from our lab (5LO2[132]); and *via* the computational design of constrained peptide (2ND2[133]) and combined computational and selection studies (5UOI[134]) from the Baker lab. Indeed today, *miniprotein design* and, broader still, Sam Gellman's *foldamer* concept[135] are research fields in themselves with considerable potential to deliver bioactive functional peptides.[136]

Coiled-coil assemblies as a special and particularly accessible case

The rational approach to protein design and engineering has led to a high level of understanding for one class of protein structure in particular; namely, the α-helical coiled coils.[137,138] In these structures, amphipathic α helices are usually encoded by 'heptad' sequence repeats of hydrophobic (*h*) and polar (*p*) residues, $(hpphppp)_{3-5}$. Two or more such helices combine through their hydrophobic faces to form rope-like bundles. Moreover, the side-chain packing—which is key to success in protein design—is intimate and highly defined through so-called knobs-into-holes (KIH) interactions.[111,139] In the late 1980s and early 1990s, the aforementioned leucine zipper (2ZTA[86]) provided the archetypal example of coiled coils, and many researchers have used it as model for experimental and computational studies of protein folding, engineering and design. Chief amongst the proponents were Kim and, his collaborator, the late and much-missed Tom Alber.[69,86,87] As part of their programme, Harbury showed that different oligomers of leucine zipper could be made simply by altering the two *h* positions of the coiled-coil heptad repeat to different combinations of leucine and isoleucine (1GCL and 1GCM[87]). I do view this as step-change in the field, as it showed that subtly different sequence repeats direct different core packings that result in different quaternary structures. Moreover, for the first time, these differences could be fully understood to deliver clear sequence-to-structure relationships for peptide assembly and rules for rational peptide design.

Over the following two decades, more rules emerged, making coiled coils probably the best understood of peptide assemblies and protein folds.[36] In turn, this has led to considerable success in the *de novo* design of coiled coils.[35,37,140] To highlight this, in Figure 2 I have labelled with 'CC' those *de novo* designed structures that test positive for KIH packing. Some of these mimic natural structures, which are mostly dimers, trimers and tetramers: 1COS,[42] 1RH4,[69] 1BB1,[38,39] 1G6U,[141] 1U0I,[43] 6FES,[99] and 6Q5S[142]; see also the rationally designed coiled-coil basis set of homomeric coiled coils from our group (4DZM and 4DZL[143] and 3RA4[144]). Others are entirely new higher-order barrel-like structures with accessible central channels that can be functionalised, embedded in membranes, or made to switch state: 3R3K,[144] 4PN8 and 4PNA,[145] 4UOT,[116] 5EHB,[146] 5EZC,[14] 6EIZ,[147] 6B85,[107] 6G67,[97] 6MCT,[108] 6YB1,[110] and 6ZT1.[148] *N.B.* Some of these examples—4PN8 and 4PNA,[145] 4UOT,[116] 5EHB,[146] 6FES,[99] 6B85[107]—are computational designs, which will be described below.

Because of this progress in rational and computational coiled-coil design, Korendovych and DeGrado view it as a solved problem.[7] I agree in two respects: First, we understand the basic physics of coiled-coil assembly sufficiently to allow the construction of accurate and realistic *in silico* models for any coiled-coil assembly. Indeed, there are a number of web-based tools that make such modelling accessible to all users.[70,149–151] Second, the basic chemistry, that is, the sequence-to-structure relationships or rules for coiled-coil design are largely complete.[7,35,37,97,142,152] However, in two other respects I would disagree: First, we do not understand the deeper chemical physics of coiled-coil stability enough to determine and compare internal energies *in silico*, let alone to calculate free energies of assembly, $\Delta G_{assembly}$, for different design targets. This will require a better and quantitative understanding of the non-covalent interactions that stabilise folded proteins, and for these to be incorporated into more-sophisticated computational forcefields. Second, we do not understand the biochemical subtleties of either natural or synthetic coiled-coil assemblies; *i.e.*, their dynamics and any conformational changes that they undergo, and how these impact on function[153,154] or can be exploited in design. That said, *de novo* coiled coils that switch state are being designed and increasingly being resolved to high resolution, see: 5EHB,[146] 6MSQ,[155] 6Q5S,[142] 7JH5,[156] and 6ZT1.[148] I will expand on both the current under-

standing and future challenges of coiled-coil assembly and design in another perspective elsewhere.

Before closing this subsection, it would be remiss of me not to mention that coiled coils and 4-helix bundles are proving to be superb scaffolds for introducing metal-based functionality into *de novo* scaffolds. I will not review this broad, detailed and growing field here and refer you to excellent reviews by others. This area has been pioneered by DeGrado,[7] Angela Lombardi,[157] Pecoraro and others, and is being advanced by them and others such as Anderson,[17] Brian Kuhlman and Hilvert,[15,96] Anna Peacock.[158]

Revolution in protein design: The emergence and establishment of computational approaches

Computational approaches played a key role in the development of rational protein design. For example, although Harbury's breakthrough relating coiled-coil sequence repeats to quaternary structures was experimental it was founded on computational work began for his PhD thesis.[87] Furthermore, the rules that emerged from this work were confirmed through computational analysis of large numbers of natural sequences,[159,160] and have been used as the cornerstones of coiled-coil design since.[35,37] Moreover, *consensus protein design*[52,56] is essentially based in computational bioinformatics with the *design rules* emanating from multiple-sequence alignments of the targeted structure (Box 1). Although these applications of computers played a key role in shifting from back-of-the-envelope (or *in biro*) designs, neither would be called computational protein design as we know it today.

As defined in Box 1, computational protein design refers to the process of setting up an *in silico* model of a target backbone structure and finding protein sequences that are compatible with it. Such approaches started to emerge in the 1990s with the realisations that: not all targets would be accessible through rational design; more-general design methods would be needed; and protein sequence and structural space is massive. My personal view is that this resonated with themes in protein engineering at the time, which were moving from making specific point mutations, through saturation mutagenesis, and into selecting 'winners' (*i.e.*, stable or functional variants) from either targeted or randomly generated mutant libraries of natural proteins.[161,162] In short, a move to *in silico* screening of large protein libraries simply made sense. The question was: how should this best be done?

Exploring and assessing impossibly large sequence and structural spaces: Broadly speaking, in the early days of computational protein design the problem was broken down into three steps: first design or choose a backbone; second, superimpose on this many side-chain combinations both in terms of sequences and conformations; and third, assess which combinations best fit the target. The assumption was that the best fits would stabilise the targeted fold when made experimentally. Thus, this was very much a process of positive design. The early approaches naturally focused on core-directed design; *i.e.*, finding combinations of hydrophobic side chains that best filled the 3D void bounded by the target backbone.[68,162,163] Mayo's contributions are particularly noteworthy here—see 1FSD.[164] Handel, Harbury and others also contributed to the development of these approaches.[69,71] To help solve the problems of sorting through many residue combinations and the impossibly large numbers of associated side-chain conformations, Bassil Dahiyat and Mayo applied branch-pruning or dead-end-elimination algorithms.[165] In regions, or branches, of design space where side-chain conformations were modelled to be incompatible with the target—for instance, where a particular residue at a specific position would always clash with the backbone—would be eliminated, or pruned, early in the searches as they would contribute unfavourable energies to any model. Since then, the building and optimisation of core packing through guided searches has been a key aspect of computational protein design.[12,116,145,166]

Allowing more backbone variety and flexibility: Early in computational design, it was apparent that using fixed or natural backbones would be limiting, and that the notion that backbones wouldn't relax or shift was naïve. Some protein tertiary and quaternary structures, such as the coiled coils and repetitive helical or solenoid protein designs can be described by a small number of geometric parameters.[36,167] In turn, parametric models for such targets can be set up to explore a swathe of tertiary and quaternary structures systematically. This has led to several successful parametric-design or partially parametric-design approaches. The first of these was Harbury's right-handed coiled-coil tetramer, which also incorporates non-proteinogenic side chains, 1RH4.[69] As noted above, the all-atom parametric modelling and design of coiled-coil structures has now advanced to a point where it is extremely robust, reliably and accessible,[70,78,116,145,151,168] and even considered a largely solved problem.[7] The confidence that coiled-coil backbones at least can be considered as defined, allows the computational design process to focus on other aspects of the problem. For example, the heterotrimeric coiled coil 1BB1[39] was achieved by optimising electrostatic interactions between different patterns of charged side chains in three chains chosen from > 2.7 million combinations of 256 possible sequences.[38]

More recent examples of what might be generally considered parametric designs have tackled: α-helical bundles, 4TQL,[116] α-helical barrels,

4PN8,[145] α/α toroids, 4YXX,[169] α-helical repeats or solenoids, 5CWB[170] and 4GMR[171]; and an α/β TIM-like barrel, 5BVL.[172] Ideally to solve the backbone aspect of the protein-design problem, all protein structures would be described parametrically. Generally, however, this is not possible—for example, see the all-β-structured 6CZH[115]—and other methods for generating backbones are needed.

Baker's fragment-based design approach, which was initially developed to address the protein-folding problem and the CASP challenge,[173] was another clear landmark in computational protein design. This has developed into the Rosetta suite of computational modelling and design tools and is being applied widely, successfully and impressively by Baker's own group, former group members, and others.[12,74,75] The first fragment-based design was Kuhlman and Baker's TOP7 design in 2003 – 1QYS.[73] In essence, the targeted backbone is pieced together from fragments from known protein structures deposited in the PDB.[20] Though there is potential for generating *de novo* proteins with similarity or homology to natural proteins, there is no doubt that this has had a huge impact on the field of computational *de novo* protein design and the problem of generating initial backbones. Since then, and through generalised frameworks such as RosettaRemodel,[174] fragment-based design has developed into one of the most successful approaches in *de novo* computational protein design. These developments are described expertly by recent reviews from Baker and DeGrado, so I will not dwell on them here.[5,7] The general approach has also inspired or otherwise spawned a range of approaches for backbone generation, including: TERMS (tertiary structural motifs) from Gevorg Grigoryan,[175] and its application to the design of *de novo* peptide-protein interactions with Amy Keating[176]; SEWING from Kuhlman and co-workers, which combines pieces of natural protein structures, *e.g.* 2N8I;[177] loop-helix-loop unit combinatorial sampling, or LUCS, from Kortemme's group, *e.g.* 6VGA;[178] and TopoBuilder from Bruno Correia and colleagues for incorporating functional elements into *de novo* protein frameworks.[103,105]

Gaps and remaining challenges: These advances in computational *de novo* protein design are allowing increasingly challenging design targets to be addressed such as structures rich in β structure – 2KL8,[179] 3WW7,[180] 5BVL,[172] and 5KPE[181]; and membrane-spanning peptide and protein assemblies – 2MUZ,[106] 6B85,[107] 6MCT,[108] 6M6Z,[84] 6X9Z,[109] and 6YB1.[110] In addition, and particularly in the last half-decade, emphasis has moved to functional protein design. For instance, designer *de novo* peptides and proteins that bind natural proteins of interest are being realised – 4OYD,[100] 5VID,[101] 6IWB,[102] 6XXV,[103] 7JH5,[156] 7JZL,[104] and 6YWC.[105] Also, with easy access to some on-line computational design tools,[151,182] the field is opening up to non-specialist users and designers – 6MRR.[183] Nonetheless, considerable challenges remain for computational design to tackle, and indeed for protein design in general.

For example, it is increasingly appreciated that computational protein design is not completely robust, and often has to be used in combination with experimental methods to deliver the goods. In other words, many of the current computational design methods get you so far, but then the experimental cavalry is needed. One example of this is in delivering stably folded structures – 5UOI,[134] and 6XXV.[103] In these examples, computational protein design delivers a library of potential designs, genes for these are then made by parallel DNA synthesis, and stable variants are selected via yeast surface display and protease selection. Interestingly, protease-based selection was envisaged and demonstrated at the aforementioned confluence of the protein-engineering and protein-design fields.[184–187] It may be that such alliances between computational and experimental approaches will suffice to deliver robust and useful designs. However, there is a clear need to improve design rules, computational methods, and particularly the forcefields used to direct the *in silico* design process and assess the designs. These would advance computational protein design further to predict whether a *de novo* protein will be stably folded or not before it is made and characterised experimentally.

Related to this, achieving small-molecule binding and catalysis comparable to that observed in nature using completely *de novo* designed scaffolds is an ongoing and difficult challenge for the field. Again, experimental screening and selection from computationally directed libraries will likely continue to be used profitably.[8] However, improving forcefields even further to capture more accurately the weak non-covalent forces that drive and discriminate protein-metal, protein-protein and protein-small molecule interactions will be critical for delivering better and more widely applicable computational protein design. Nonetheless, progress is being made in the recognition and tight binding of small molecules by *de novo* peptides and proteins – 5TGY,[188] 6CZH,[115] 6EIZ,[147] and 6W70.[189] In addition, catalytic activities, with in some cases very impressive catalytic efficiencies, are being achieved by porting catalytic triads and metals into completely *de novo* frameworks,[17,190] with structures being resolved for some of these – 5EZC,[14] 5UGK,[191] and 5OD1.[15]

Aligned with improving binding and catalysis by *de novo* proteins, there is the challenge of incorporating dynamics and conformational changes in protein design. Nature evolves proteins to be thermodynamically stable up to thresholds set by the prevailing conditions; there is no pressure for her to do anything more. Indeed, there are clear advantages of limiting protein stability, not least the ability to turnover

and recycle protein chains once they have served their purpose. By contrast, protein designers have always prided themselves on being able to achieve hyperstable proteins.[33,116,145,188] This might help raise the thermodynamic bar of what is possible in folding polypeptide chains to equal or surpass that set by proteins from thermophiles.[192] However, it may bring with it problems for functional design. For in-cell applications, the aforementioned protein turnover will be required. More generally, it is well established that dynamics, conformational changes, switching and allostery—which may be hampered by hyperstability—are all critical for natural protein functions and regulation. This has always been on the mind of protein designers,[40,47,193–197] and, as noted above, structural transitions are being incorporated into coiled-coil designs – 5EHB,[146] 6MSQ,[155] 6Q5S,[142] 7JH5,[156] 6ZT1,[148] and.[198] General approaches to this *multistate design* problem are being developed and tested, though currently on natural systems. For example, Roberto Chica is developing dynamic and native conformational exchangers (DANCERS)[199]; Patrick Barth is engineering allosteric microswitches into transmembrane receptors[200]; and Sophie Barbe and Thomas Schiex are using advanced computing for positive multistate protein design.[201] I anticipate that these early examples and developing methods will help make inroads in the design of synthetic proteins to increasingly match the sophistication and wonder of natural proteins.

Conclusion

In summary, the past three decades have witnessed considerable and impressive growth and development of *de novo* protein design. The field has moved on from largely empirical and minimalist approaches that test our basic understanding of protein folding, through rational approaches that develop and apply sequence-to-structure relationships or rules for protein design, and onto computational protein design, which is delivering complex protein structures and functions. Along with this, more *de novo* protein structures are being resolved to high resolution (see Figures 1 and 2) and, increasingly, these match the design models and functional expectations. This has delivered > 100 structures that can be fed back into the design-build-test-learn cycle to improve protein design methods and outcomes in future. That's progress, and the future for *de novo* protein design looks very bright indeed.

Nonetheless, and as I have started to outline above, there are challenges ahead. These include the provision of better parametric, fragment-based, and other methods to generate the initial backbones for *de novo* design. These would allow the further exploration of the dark matter of protein space,[53,54] and for *de novo* proteins to be built around targeted functions rather than grafting functions onto pre-existing scaffolds.[18,105] Also, whilst there is a growing number of *de novo* proteins that bind small molecules and/or catalyse chemical reactions, the routine and accurate design of specific and tight binding, and of catalysis remains challenging. Related to these functional targets, our understanding of protein dynamics and conformational changes and our ability to capture these in *de novo* proteins remains rudimentary. All of these challenges must be addressed if we are to claim that *de novo* protein design has come of age,[5] and that it can operate alongside protein engineering. Most likely, this will require both advances in computational design methods—for instance, incorporating machine learning[202] and virtual reality—and improving our fundamental understanding of the chemical physics of protein structure and function to deliver better forcefields for guiding and assessing protein designs.[5,12,203]

Finally, given the progress on and power of computational protein design, we should ask what is the value of continuing with other approaches; namely, minimal and rational design. My view is that the case for pursuing purely minimal design is increasingly difficult to make. Though, respected colleagues of mine do present clear arguments for continuing this for specific targets,[6,204,205] and it has the advantage of throwing up interesting surprises where targeted rational and computational approaches may not. The case for continuing rational design is more balanced, which I illustrate with recent examples from the design of α-helical barrels. These are reasonably rare in nature,[206] and engineering and design routes into this space have been discovered serendipitously[144,207] – 3R3K. Following this, *de novo* α-helical barrels have been achieved through computational designs by the Baker lab and my own group[116,145] – 4PN8, 4PNA and 4UOT. Subsequent rational and empirical experimental studies of these have shown that these are not all robust, with point or permutants to both sets causing topological rearrangements or even complete collapse of the structures – 6G6F and 6G6H.[97] Consequently, the empirical work has led to more-robust design rules not apparent from the computational designs. In turn, these design principles and examples of water-soluble α-helical barrels have been used to tackle the more-challenging problem of specifying helix-helix interactions in membranes by DeGrado (6MCT[108]), and to design membrane-spanning ion-channel proteins and peptides from the Baker lab (6M6Z[107]) and our group (6YB1[110]). In this vein, I hope and anticipate that the rational rules-based and powerful computational approaches to design can be combined further and more widely to achieve better *de novo* protein designs. By better I mean that we can understand and rationalise the designs delivered, and that we can use them for

applications in biotechnology, medicine and synthetic biology. Encouragingly, this is already happening with, for example, protein-design methods being applied in biosensing.[147,208–210]

CRediT authorship contribution statement

Derek N. Woolfson: Conceptualization, Investigation, Funding acquisition.

Acknowledgements

I should like to thank Prasun Kumar, and Ilyas Grandguillaume and Bernard Offmann for help in searching the PDB for the *de novo* designed protein structures presented in Figures 1 and 2; Emily Baker, Will Dawson, Luigi Dicostanzo, Elise Naudin, and Fabio Pirro for extremely insightful comments on this manuscript; past and present members of my research group, undergraduate students, and colleagues in biochemistry and chemistry at the University of Bristol for helping me form some of the ideas presented here; and the BBSRC (BB/L01386X/1) and the University of Bristol for financial support for BrisSynBio and the Bristol BioDesign Institute.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received 6 June 2021;
Accepted 12 July 2021;
Available online xxxx

Keywords:
chemical biology;
peptide self-assembly;
protein engineering;
protein folding;
and synthetic biology

References

1. https://www.proteinsociety.org/page/meetings
2. Bryson, J.W., Betz, S.F., Lu, H.S., Suich, D.J., Zhou, H.X. X., Oneil, K.T., et al., (1995). Protein design - a hierarchical approach. *Science*, **270**, 935–941.
3. Hill, R.B., Raleigh, D.P., Lombardi, A., Degrado, W.F., (2000). De novo design of helical bundles as models for understanding protein folding and function. *Accounts Chem. Res.*, **33**, 745–754.
4. Regan, L., Caballero, D., Hinrichsen, M.R., Virrueta, A., Williams, D.M., O'Hern, C.S., (2015). Protein design: past, present, and future. *Biopolymers*, **104**, 334–350.
5. Huang, P.S., Boyken, S.E., Baker, D., (2016). The coming of age of de novo protein design. *Nature*, **537**, 320–327.
6. Curnow, P., (2019). Designing minimalist membrane proteins. *Biochem. Soc. T.*, **47**, 1233–1245.
7. Korendovych, I.V., DeGrado, W.F., (2020). De novo protein design, a retrospective. *Q. Rev. Biophys.*, **53**
8. Zeymer, C., Hilvert, D., (2018). Directed evolution of protein catalysts. *Annu. Rev. Biochem.*, **87**, 131–157.
9. Yang, K.K., Wu, Z., Arnold, F.H., (2019). Machine-learning-guided directed evolution for protein engineering. *Nat. Methods*, **16**, 687–694.
10. https://www.nobelprize.org/prizes/chemistry/2018/summary/.
11. Alford, R.F., Leaver-Fay, A., Jeliazkov, J.R., O'Meara, M. J., DiMaio, F.P., Park, H., et al., (2017). The rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.*, **13**, 3031–3048.
12. Leman, J.K., Weitzner, B.D., Lewis, S.M., Adolf-Bryfogle, J., Alam, N., Alford, R.F., et al., (2020). Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat. Methods*, **17**, 665–680.
13. Goldenzweig, A., Goldsmith, M., Hill, S.E., Gertman, O., Laurino, P., Ashani, Y., et al., (2016). Automated structure- and sequence-based design of proteins for high bacterial expression and stability. *Mol. Cell*, **63**, 337–346.
14. Burton, A.J., Thomson, A.R., Dawson, W.M., Brady, R.L., Woolfson, D.N., (2016). Installing hydrolytic activity into a completely de novo protein framework. *Nat. Chem.*, **8**, 837–844.
15. Studer, S., Hansen, D.A., Pianowski, Z.L., Mittl, P.R.E., Debon, A., Guffy, S.L., et al., (2018). Evolution of a highly active and enantiospecific metalloenzyme from short peptides. *Science*,. 362:1285-+..
16. Kries, H., Blomberg, R., Hilvert, D., (2013). De novo enzymes by computational design. *Curr. Opin. Chem. Biol.*, **17**, 221–228.
17. Bunzel, H.A., Anderson, J.L.R., Mulholland, A.J., (2021). Designing better enzymes: Insights from directed evolution. *Curr. Opin. Struc. Biol.*, **67**, 212–218.
18. Dawson, W.M., Rhys, G.G., Woolfson, D.N., (2019). Towards functional de novo designed proteins. *Curr. Opin. Chem. Biol.*, **52**, 102–111.
19. Yue, K., Dill, K.A., (1992). Inverse protein folding problem - designing polymer sequences. *P. Natl. Acad. Sci. U.S. A.*, **89**, 4163–4167.
20. Burley, S.K., Bhikadiya, C., Bi, C.X., Bittrich, S., Chen, L., Crichlow, G.V., et al., (2021). RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.*, **49**, D437–D451.
21. Degrado, W.F., Wasserman, Z.R., Lear, J.D., (1989). Protein design, a minimalist approach. *Science*, **243**, 622–628.
22. Kamtekar, S., Schiffer, J.M., Xiong, H.Y., Babik, J.M., Hecht, M.H., (1993). Protein design by binary patterning of polar and nonpolar amino-acids. *Science*, **262**, 1680–1685.

23. Hecht, M.H., Das, A., Go, A., Bradley, L.H., Wei, Y.N., (2004). De novo proteins from designed combinatorial libraries. *Protein Sci.*, **13**, 1711–1723.
24. Richardson, J.S., Richardson, D.C., (1989). The de novo design of protein structures. *Trends Biochem. Sci.*, **14**, 304–309.
25. Hecht, M.H., Richardson, J.S., Richardson, D.C., Ogden, R.C., (1990). De novo design, expression, and characterization of felix - a 4-helix bundle protein of native-like sequence. *Science*, **249**, 884–891.
26. Quinn, T.P., Tweedy, N.B., Williams, R.W., Richardson, J. S., Richardson, D.C., (1994). Betadoublet - De-novo design, synthesis, and characterization of a beta-sandwich protein. *P. Natl. Acad. Sci. U.S.A.*, **91**, 8747–8751.
27. Koder, R.L., Anderson, J.L.R., Solomon, L.A., Reddy, K. S., Moser, C.C., Dutton, P.L., (2009). Design and engineering of an O-2 transport protein. *Nature*, **458**, 305–U64.
28. Farid, T.A., Kodali, G., Solomon, L.A., Lichtenstein, B.R., Sheehan, M.M., Fry, B.A., et al., (2013). Elementary tetrahelical protein design for diverse oxidoreductase functions. *Nat. Chem. Biol.*,. 9:826-+..
29. Hodges, R.S., Saund, A.K., Chong, P.C.S., Stpierre, S.A., Reid, R.E., (1981). Synthetic model for 2-stranded alpha-helical coiled-coils - design, synthesis, and characterization of an 86-residue analog of tropomyosin. *J. Biol. Chem.*, **256**, 1214–1224.
30. Hodges, R.S., (1996). De novo design of alpha-helical proteins: Basic research to medical applications. *Biochem. Cell Biol.*, **74**, 133–154.
31. https://www.americanpeptidesociety.org/obituary/hodges-robert/.
32. Nanda, V., Koder, R.L., (2010). Designing artificial enzymes by intuition and computation. *Nat. Chem.*, **2**, 15–24.
33. Regan, L., Degrado, W.F., (1988). Characterization of a helical protein designed from 1st principles. *Science*, **241**, 976–978.
34. Regan, L., Clarke, N.D., (1990). A tetrahedral zinc(Ii)-binding site introduced into a designed protein. *Biochemistry-Us.*, **29**, 10878–10883.
35. Woolfson, D.N., (2005). The design of coiled-coil structures and assemblies. *Fibrous Proteins: Coiled-Coils, Collagen and Elastomers.*,. 70:79-+..
36. Lupas, A.N., Bassler, J., (2017). Coiled coils - a model system for the 21st century. *Trends Biochem. Sci.*, **42**, 130–140.
37. Woolfson, D.N., (2017). Coiled-coil design: updated and upgraded. *Subcell. Biochem.*, **82**, 35–61.
38. Nautiyal, S., Woolfson, D.N., King, D.S., Alber, T., (1995). A designed heterotrimeric coiled-coil. *Biochemistry-Us.*, **34**, 11645–11651.
39. Nautiyal, S., Alber, T., (1999). Crystal structure of a designed, thermostable; heterotrimeric coiled coil. *Protein Sci.*, **8**, 84–90.
40. Zimenkov, Y., Dublin, S.N., Ni, R., Tu, R.S., Breedveld, V., Apkarian, R.P., et al., (2006). Rational design of a reversible pH-responsive switch for peptide self-assembly. *J. Am. Chem. Soc.*, **128**, 6770–6771.
41. Rele, S., Song, Y.H., Apkarian, R.P., Qu, Z., Conticello, V. P., Chaikof, E.L., (2007). D-periodic collagen-mimetic microfibers. *J. Am. Chem. Soc.*, **129**, 14780–14787.
42. Lovejoy, B., Choe, S., Cascio, D., Mcrorie, D.K., Degrado, W.F., Eisenberg, D., (1993). Crystal-structure of a synthetic triple-stranded alpha-helical bundle. *Science*, **259**, 1288–1293.
43. Lindhout, D.A., Litowski, J.R., Mercier, P., Hodges, R.S., Sykes, B.D., (2004). NMR solution structure of a highly stable de novo heterodimeric coiled-coil. *Biopolymers*, **75**, 367–375.
44. O'Shea, E.K., Lumb, K.J., Kim, P.S., (1993). Peptide velcro - design of a heterodimeric coiled-coil. *Curr. Biol.*, **3**, 658–667.
45. Yu, F.T., Cangelosi, V.M., Zastrow, M.L., Tegoni, M., Plegaria, J.S., Tebo, A.G., et al., (2014). Protein design: toward functional metalloenzymes. *Chem. Rev.*, **114**, 3495–3578.
46. Pandya, M.J., Spooner, G.M., Sunde, M., Thorpe, J.R., Rodger, A., Woolfson, D.N., (2000). Sticky-end assembly of a designed peptide fiber provides insight into protein fibrillogenesis. *Biochemistry-Us.*, **39**, 8728–8734.
47. Ciani, B., Hutchinson, E.G., Sessions, R.B., Woolfson, D. N., (2002). A designed system for assessing how sequence affects alpha to beta conformational transitions in proteins. *J. Biol. Chem.*, **277**, 10150–10155.
48. Banwell, E.F., Abelardo, E.S., Adams, D.J., Birchall, M.A., Corrigan, A., Donald, A.M., et al., (2009). Rational design and application of responsive alpha-helical peptide hydrogels. *Nat. Mater.*, **8**, 596–600.
49. Kortemme, T., Ramirez-Alvarado, M., Serrano, L., (1998). Design of a 20-amino acid, three-stranded beta-sheet protein. *Science*, **281**, 253–256.
50. Main, E.R.G., Jackson, S.E., Regan, L., (2003). The folding and design of repeat proteins: reaching a consensus. *Curr. Opin. Struc. Biol.*, **13**, 482–489.
51. Mosavi, L.K., Cammett, T.J., Desrosiers, D.C., Peng, Z. Y., (2004). The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci.*, **13**, 1435–1448.
52. Porebski, B.T., Buckle, A.M., (2016). Consensus protein design. *Protein Eng. Des. Sel.*, **29**, 245–251.
53. Taylor, W.R., Chelliah, V., Hollup, S.M., MacDonald, J.T., Jonassen, I., (2009). Probing the "dark matter" of protein fold space. *Structure*, **17**, 1244–1252.
54. Woolfson, D.N., Bartlett, G.J., Burton, A.J., Heal, J.W., Niitsu, A., Thomson, A.R., et al., (2015). De novo protein design: how do we expand into the universe of possible protein structures?. *Curr. Opin. Struc. Biol.*, **33**, 16–26.
55. Jackel, C., Bloom, J.D., Kast, P., Arnold, F.H., Hilvert, D., (2010). Consensus protein design without phylogenetic bias. *J. Mol. Biol.*, **399**, 541–546.
56. Pluckthun, A., (2015). Designed ankyrin repeat proteins (DARPins): Binding proteins for research, diagnostics, and therapy. *Annu. Rev. Pharmacol.*, **55**, 489–511.
57. Krizek, B.A., Amann, B.T., Kilfoil, V.J., Merkle, D.L., Berg, J.M., (1991). A consensus zinc finger peptide - design, high-affinity metal-binding, a ph-dependent structure, and a his to cys sequence variant. *J. Am. Chem. Soc.*, **113**, 4518–4523.
58. Kim, C.A., Berg, J.M., (1996). A 2.2 angstrom resolution crystal structure of a designed zinc finger protein bound to DNA. *Nat. Struct. Biol.*, **3**, 940–945.
59. Struthers, M.D., Cheng, R.P., Imperiali, B., (1996). Design of a monomeric 23-residue polypeptide with defined tertiary structure. *Science*, **271**, 342–345.

60. Mosavi, L.K., Minor, D.L., Peng, Z.Y., (2002). Consensus-derived structural determinants of the ankyrin repeat motif. *P. Natl. Acad. Sci. U.S.A.*, **99**, 16029–16034.
61. Kohl, A., Binz, H.K., Forrer, P., Stumpp, M.T., Pluckthun, A., Grutter, M.G., (2003). Designed to be stable: Crystal structure of a consensus ankyrin repeat protein. *P. Natl. Acad. Sci. U.S.A.*, **100**, 1700–1705.
62. Main, E.R.G., Xiong, Y., Cocco, M.J., D'Andrea, L., Regan, L., (2003). Design of stable alpha-helical arrays from an idealized TPR motif. *Structure*, **11**, 497–508.
63. Porebski, B.T., Nickson, A.A., Hoke, D.E., Hunter, M.R., Zhu, L.G., McGowan, S., et al., (2015). Structural and dynamic properties that govern the stability of an engineered fibronectin type III domain. *Protein Eng. Des. Sel.*, **28**, 67–78.
64. Lehmann, M., Lopez-Ulibarri, R., Loch, C., Viarouge, C., Wyss, M., Van Loon, A.P.G.M., (2000). Exchanging the active site between phytases for altering the functional properties of the enzyme. *Protein Sci.*, **9**, 1866–1872.
65. Urvoas, A., Guellouz, A., Valerio-Lepiniec, M., Graille, M., Durand, D., Desravines, D.C., et al., (2010). Design, production and molecular structure of a new family of artificial alpha-helicoidal repeat proteins (alpha rep) based on thermostable HEAT-like repeats. *J. Mol. Biol.*, **404**, 307–327.
66. Urvoas, A., Valerio-Lepiniec, M., Minard, P., (2012). Artificial proteins from combinatorial approaches. *Trends Biotechnol.*, **30**, 512–520.
67. Coquille, S., Filipovska, A., Chia, T., Rajappa, L., Lingford, J.P., Razif, M.F.M., et al., (2014). An artificial PPR scaffold for programmable RNA recognition. *Nat. Commun.*,. 5:ARTN 5729.
68. Street, A.G., Mayo, S.L., (1999). Computational protein design. *Structure*, **7** R105-R9.
69. Harbury, P.B., Plecs, J.J., Tidor, B., Alber, T., Kim, P.S., (1998). High-resolution protein design with backbone freedom. *Science*, **282**, 1462–1467.
70. Grigoryan, G., DeGrado, W.F., (2011). Probing designability via a generalized model of helical bundle geometry. *J. Mol. Biol.*, **405**, 1079–1100.
71. Desjarlais, J.R., Handel, T.M., (1995). De-novo design of the hydrophobic cores of proteins. *Protein Sci.*, **4**, 2006–2018.
72. Dahiyat, B.I., Mayo, S.L., (1997). De novo protein design: Fully automated sequence selection. *Science*, **278**, 82–87.
73. Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L., Baker, D., (2003). Design of a novel globular protein fold with atomic-level accuracy. *Science*, **302**, 1364–1368.
74. Leman, J.K., Weitzner, B.D., Renfrew, P.D., Lewis, S.M., Moretti, R., Watkins, A.M., et al., (2020). Better together: Elements of successful scientific software development in a distributed collaborative community. *PLoS Comput. Biol.*, **16**
75. Meinen, B., Bahl, C.D., (2021). Breakthroughs in computational design methods open up new frontiers for de novo protein engineering. *Protein Eng. Des. Sel.*, **34**
76. Gainza, P., Roberts, K.E., Georgiev, I., Lilien, R.H., Keedy, D.A., Chen, C.Y., et al., (2013). OSPREY: protein design with ensembles, flexibility, and provable algorithms. *Methods in Protein Design*, **523**, 87–107.
77. Hallen, M.A., Martin, J.W., Ojewole, A., Jou, J.D., Lowegard, A.U., Frenkel, M.S., et al., (2018). OSPREY 3.0: Open-source protein redesign for you, with powerful new features. *J. Comput. Chem.*, **39**, 2494–2507.
78. Wood, C.W., Heal, J.W., Thomson, A.R., Bartlett, G.J., Ibarra, A.A., Brady, R.L., et al., (2017). ISAMBARD: an open-source computational environment for biomolecular analysis, modelling and design. *Bioinformatics*, **33**, 3043–3050.
79. Creighton, T.E., (1990). Protein folding. *Biochem. J.*, **270**, 1–16.
80. Akerfeldt, K.S., Lear, J.D., Wasserman, Z.R., Chung, L. A., Degrado, W.F., (1993). Synthetic peptides as models for ion channel proteins. *Accounts Chem. Res.*, **26**, 191–197.
81. Johnsson, K., Allemann, R.K., Widmer, H., Benner, S.A., (1993). Synthesis, structure and activity of artificial, rationally designed catalytic polypeptides. *Nature*, **365**, 530–532.
82. DeGrado, W.F., Summa, C.M., Pavone, V., Nastri, F., Lombardi, A., (1999). De novo design and structural characterization of proteins and metalloproteins. *Annu. Rev. Biochem.*, **68**, 779–819.
83. Hughes, S.A., Wang, F.B., Wang, S.Y., Kreutzberger, M. A.B., Osinski, T., Orlova, A., et al., (2019). Ambidextrous helical nanotubes from self-assembly of designed helical hairpin motifs. *P. Natl.. Acad. Sci. U.S.A.*, **116**, 14456–14464.
84. Xu, C.F., Lu, P.L., El-Din, T.M.G., Pei, X.Y., Johnson, M. C., Uyeda, A., et al., (2020). Computational design of transmembrane pores. *Nature*,. 585:129-+..
85. Wang, F.B., Gnewou, O., Modlin, C., Beltran, L.C., Xu, C. F., Su, Z.L., et al., (2021). Structural analysis of cross alpha-helical nanotubes provides insight into the designability of filamentous peptide nanomaterials. *Nat. Commun.*,. 12:ARTN 407.
86. O'Shea, E.K., Klemm, J.D., Kim, P.S., Alber, T., (1991). X-ray structure of the Gcn4 leucine zipper, a 2-stranded, parallel coiled coil. *Science*, **254**, 539–544.
87. Harbury, P.B., Zhang, T., Kim, P.S., Alber, T., (1993). A switch between 2-stranded, 3-stranded and 4-stranded coiled coils in Gcn4 leucine-zipper mutants. *Science*, **262**, 1401–1407.
88. Ottesen, J.J., Imperiali, B., (2001). Design of a discretely folded mini-protein motif with predominantly beta-structure. *Nat. Struct. Biol.*, **8**, 535–539.
89. Neidigh, J.W., Fesinmeyer, R.M., Andersen, N.H., (2002). Designing a 20-residue protein. *Nat. Struct. Biol.*, **9**, 425–430.
90. Nicoll, A.J., Allemann, R.K., (2004). Nucleophilic and general acid catalysis at physiological pH by a designed miniature esterase. *Org. Biomol. Chem.*, **2**, 2175–2780.
91. de Bono, S., Riechmann, L., Girard, E., Williams, R.L., Winter, G., (2005). A segment of cold shock protein directs the folding of a combinatorial protein. *P. Natl. Acad. Sci. U.S.A.*, **102**, 1396–1401.
92. Mansy, S.S., Zhang, J.L., Kummerle, R., Nilsson, M., Chou, J.J., Szostak, J.W., et al., (2007). Structure and evolutionary analysis of a non-biological ATP-binding protein. *J. Mol. Biol.*, **371**, 501–513.
93. King, N.P., Sheffler, W., Sawaya, M.R., Vollmar, B.S., Sumida, J.P., Andre, I., et al., (2012). Computational

design of self-assembling protein nanomaterials with atomic level accuracy. *Science*, **336**, 1171–1174.
94. Longo, L.M., Lee, J., Blaber, M., (2013). Simplified protein design biased for prebiotic amino acids yields a foldable, halophilic protein. *P. Natl. Acad. Sci. U.S.A.*, **110**, 2135–2139.
95. Zhou, L., Bosscher, M., Zhang, C.S., Ozcubukcu, S., Zhang, L., Zhang, W., et al., (2014). A protein engineered to bind uranyl selectively and with femtomolar affinity. *Nat. Chem.*, **6**, 236–241.
96. Der, B.S., Machius, M., Miley, M.J., Mills, J.L., Szyperski, T., Kuhlman, B., (2012). Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. *J. Am. Chem. Soc.*, **134**, 375–385.
97. Rhys, G.G., Wood, C.W., Lang, E.J.M., Mulholland, A.J., Brady, R.L., Thomson, A.R., et al., (2018). Maintaining and breaking symmetry in homomeric coiled-coil assemblies. *Nat. Commun.*,. 9:ARTN 4132.
98. Walsh, S.T.R., Cheng, H., Bryson, J.W., Roder, H., DeGrado, W.F., (1999). Solution structure and dynamics of a de novo designed three-helix bundle protein. *P. Natl. Acad. Sci. U.S.A.*, **96**, 5486–5491.
99. ElGamacy, M., Coles, M., Ernst, P., Zhu, H.B., Hartmann, M.D., Pluckthun, A., et al., (2018). An interface-driven design strategy yields a novel, corrugated protein architecture. *Acs Synth. Biol.*, **7**, 2226–2235.
100. Procko, E., Berguig, G.Y., Shen, B.W., Song, Y.F., Frayo, S., Convertine, A.J., et al., (2014). A computationally designed inhibitor of an epstein-barr viral Bcl-2 protein induces apoptosis in infected cells. *Cell*, **157**, 1644–1656.
101. Chevalier, A., Silva, D.A., Rocklin, G.J., Hicks, D.R., Vergara, R., Murapa, P., et al., (2017). Massively parallel de novo protein design for targeted therapeutics. *Nature*,. 550:74-+..
102. Giordano-Attianese, G., Gainza, P., Gray-Gaillard, E., Cribioli, E., Shui, S.L., Kim, S., et al., (2020). A computationally designed chimeric antigen receptor provides a small-molecule safety switch for T-cell therapy (vol 41, pg 568, 2019). *Nat. Biotechnol.*, **38** 503-.
103. Sesterhenn, F., Yang, C., Bonet, J., Cramer, J.T., Wen, X. L., Wang, Y.M., et al., (2020). De novo protein design enables the precise induction of RSV-neutralizing antibodies. *Science*,. 368:730-+..
104. Cao, L.X., Goreshnik, I., Coventry, B., Case, J.B., Miller, L., Kozodoy, L., et al., (2020). De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science*,. 370:426-+..
105. Yang, C., Sesterhenn, F., Bonet, J., van Aalen, E.A., Scheller, L., Abriata, L.A., et al., (2021). Bottom-up de novo design of functional proteins with complex structural features. *Nat. Chem. Biol.*, **17** 492 U306.
106. Joh, N.H., Wang, T., Bhate, M.P., Acharya, R., Wu, Y.B., Grabe, M., et al., (2014). De novo design of a transmembrane Zn2+-transporting four-helix bundle. *Science*, **346**, 1520–1524.
107. Lu, P.L., Min, D.Y., DiMaio, F., Wei, K.Y., Vahey, M.D., Boyken, S.E., et al., (2018). Accurate computational design of multipass transmembrane proteins. *Science*, **359**, 1042–1046.
108. Mravic, M., Thomaston, J.L., Tucker, M., Solomon, P.E., Liu, L.J., DeGrado, W.F., (2019). Packing of apolar side chains enables accurate design of highly stable membrane proteins. *Science*,. 363:1418-+.
109. Vorobieva, A.A., White, P., Liang, B.Y., Horne, J.E., Bera, A.K., Chow, C.M., et al., (2021). De novo design of transmembrane beta barrels. *Science*,. 371:801-+..
110. Scott, A.J., Niitsu, A., Kratochvil, H.T., Lang, E.J.M., Sengel, J.T., Dawson, W.M., et al., (2021). Constructing ion channels from water-soluble alpha-helical barrels. *Nat. Chem.*, **13**, 643–650.
111. Walshaw, J., Woolfson, D.N., (2001). SOCKET: A program for identifying and analysing coiled-coil motifs within protein structures. *J. Mol. Biol.*, **307**, 1427–1450.
112. Chakrabartty, A., Baldwin, R.L., (1995). Stability of alpha-helices. *Adv. Protein Chem.*, **46**, 141–176.
113. Makin, O.S., Serpell, L.C., (2005). Structures for amyloid fibrils. *FEBS J.*, **272**, 5950–5961.
114. Hughes, R.M., Waters, M.L., (2006). Model systems for beta-hairpins and beta-sheets. *Curr. Opin. Struc. Biol.*, **16**, 514–524.
115. Dou, J.Y., Vorobieva, A.A., Sheffler, W., Doyle, L.A., Park, H., Bick, M.J., et al., (2018). De novo design of a fluorescence-activating beta-barrel. *Nature*,. 561:485-+..
116. Huang, P.S., Oberdorfer, G., Xu, C.F., Pei, X.Y., Nannenga, B.L., Rogers, J.M., et al., (2014). High thermodynamic stability of parametrically designed helical bundles. *Science*, **346**, 481–485.
117. Hill, C.P., Anderson, D.H., Wesson, L., Degrado, W.F., Eisenberg, D., (1990). Crystal-structure of alpha-1 - implications for protein design. *Science*, **249**, 543–546.
118. O'Neil, K.T., Degrado, W.F., (1990). A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino-acids. *Science*, **250**, 646–651.
119. Stranges, P.B., Kuhlman, B., (2013). A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci.*, **22**, 74–82.
120. Bowie, J.U., Luthy, R., Eisenberg, D., (1991). A method to identify protein sequences that fold into a known 3-dimensional structure. *Science*, **253**, 164–170.
121. Sippl, M.J., (1993). Boltzmann principle, knowledge-based mean fields and protein-folding - an approach to the computational determination of protein structures. *J. Comput. Aid. Mol. Des.*, **7**, 473–501.
122. Eddy, S.R., (1998). Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.
123. Rohl, C.A., Strauss, C.E.M., Misura, K.M.S., Baker, D., (2004). Protein structure prediction using rosetta. *Method Enzymol.*,. 383:66-+.
124. Senior, A.W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., et al., (2020). Improved protein structure prediction using potentials from deep learning. *Nature*,. 577:706-+.
125. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10**, 845–858.
126. AlQuraishi, M., (2019). AlphaFold at CASP13. *Bioinformatics*, **35**, 4862–4865.
127. Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., Moult, J., (2019). Critical assessment of methods of protein structure prediction (CASP)-Round XIII. *Proteins*, **87**, 1011–1020.
128. Klug, A., Schwabe, J.W.R., Motifs, P., (1995). 5. Zinc fingers. *FASEB J.*, **9**, 597–604.

129. Pabo, C.O., Peisach, E., Grant, R.A., (2001). Design and selection of novel Cys(2)His(2) zinc finger proteins. *Annu. Rev. Biochem.*, **70**, 313–340.
130. Desjarlais, J.R., Berg, J.M., (1992). Toward rules relating zinc finger protein sequences and DNA-binding site preferences. *P. Natl. Acad. Sci. U.S.A.*, **89**, 7345–7349.
131. Cochran, A.G., Skelton, N.J., Starovasnik, M.A., (2001). Tryptophan zippers: Stable, monomeric beta-hairpins. *P. Natl. Acad. Sci. U.S.A.*, **98**, 5578–5583.
132. Baker, E.G., Williams, C., Hudson, K.L., Bartlett, G.J., Heal, J.W., Goff, K.L.P., et al., (2017). Engineering protein stability with atomic precision in a monomeric miniprotein. *Nat. Chem. Biol.*,. 13:764-+.
133. Bhardwaj, G., Mulligan, V.K., Bahl, C.D., Gilmore, J.M., Harvey, P.J., Cheneval, O., et al., (2016). Accurate de novo design of hyperstable constrained peptides. *Nature*,. 538:329-+.
134. Rocklin, G.J., Chidyausiku, T.M., Goreshnik, I., Ford, A., Houliston, S., Lemak, A., et al., (2017). Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science*, **357**, 168–174.
135. Gellman, S.H., (1998). Foldamers: A manifesto. *Accounts Chem. Res.*, **31**, 173–180.
136. Baker, E.G., Bartlett, G.J., Goff, K.L.P., Woolfson, D.N., (2017). Miniprotein design: past, present, and prospects. *Accounts Chem. Res.*, **50**, 2085–2092.
137. Lupas, A.N., Gruber, M., (2005). The structure of alpha-helical coiled coils. *Fibrous Proteins: Coiled-Coils, Collagen and Elastomers*,. 70:37-+.
138. Lupas, A.N., Bassler, J., Dunin-Horkawicz, S., (2017). The structure and topology of alpha-helical coiled coils. *Subcell. Biochem.*, **82**, 95–129.
139. Crick, F.H.C., (1953). The packing of alpha-helices - simple coiled-coils. *Acta Crystallogr.*, **6**, 689–697.
140. Lapenta, F., Aupic, J., Strmsek, Z., Jerala, R., (2018). Coiled coil protein origami: from modular design principles towards biotechnological applications. *Chem. Soc. Rev.*, **47**, 3530–3542.
141. Ogihara, N.L., Ghirlanda, G., Bryson, J.W., Gingery, M., DeGrado, W.F., Eisenberg, D., (2001). Design of three-dimensional domain-swapped dimers and fibrous oligomers. *P Natl Acad Sci USA*, **98**, 1404–1409.
142. Rhys, G.G., Wood, C.W., Beesley, J.L., Zaccai, N.R., Burton, A.J., Brady, R.L., et al., (2019). Navigating the structural landscape of de novo alpha-helical bundles. *J. Am. Chem. Soc.*, **141**, 8787–8797.
143. Fletcher, J.M., Boyle, A.L., Bruning, M., Bartlett, G.J., Vincent, T.L., Zaccai, N.R., et al., (2012). A basis set of de novo coiled-coil peptide oligomers for rational protein design and synthetic biology. *ACS Synth. Biol.*, **1**, 240–250.
144. Zaccai, N.R., Chi, B., Thomson, A.R., Boyle, A.L., Bartlett, G.J., Bruning, M., et al., (2011). A de novo peptide hexamer with a mutable channel. *Nat. Chem. Biol.*, **7**, 935–941.
145. Thomson, A.R., Wood, C.W., Burton, A.J., Bartlett, G.J., Sessions, R.B., Brady, R.L., et al., (2014). Computational design of water-soluble alpha-helical barrels. *Science*, **346**, 485–488.
146. Lizatovic, R., Aurelius, O., Stenstrom, O., Drakenberg, T., Akke, M., Logan, D.T., et al., (2016). A de novo designed coiled-coil peptide with a reversible pH-induced oligomerization switch. *Structure*, **24**, 946–955.
147. Thomas, F., Dawson, W.M., Lang, E.J.M., Burton, A.J., Bartlett, G.J., Rhys, G.G., et al., (2018). De novo-designed alpha-helical barrels as receptors for small molecules. *ACS Synth. Biol.*, **7**, 1808–1816.
148. Dawson, W.M., Lang, E.J.M., Rhys, G.G., Shelley, K.L., Williams, C., Brady, R.L., et al., (2021). Structural resolution of switchable states of a de novo peptide assembly. *Nat. Commun.*, **12**
149. Wood, C.W., Bruning, M., Ibarra, A.A., Bartlett, G.J., Thomson, A.R., Sessions, R.B., et al., (2014). CCBuilder: an interactive web-based tool for building, designing and assessing coiled-coil protein assemblies. *Bioinformatics*, **30**, 3029–3035.
150. Guzenko, D., Strelkov, S.V., (2018). CCFold: rapid and accurate prediction of coiled-coil structures and application to modelling intermediate filaments. *Bioinformatics*, **34**, 215–222.
151. Wood, C.W., Woolfson, D.N., (2018). CCBuilder 2.0: Powerful and accessible coiled-coil modeling. *Protein Sci.*, **27**, 103–111.
152. Dawson, W.M., Martin, F.J.O., Rhys, G.G., Shelley, K.L., Brady, R.L., Woolfson, D.N., (2021). Coiled coils 9-to-5: rational de novo design of alpha-helical barrels with tunable oligomeric states. *Chem. Sci.*, **12**, 6923–6928.
153. McNamara, C., Zinkernagel, A.S., Macheboeuf, P., Cunningham, M.W., Nizet, V., Ghosh, P., (2008). Coiled-coil irregularities and instabilities in group A Streptococcus M1 are required for virulence. *Science*, **319**, 1405–1408.
154. Hulko, M., Berndt, F., Gruber, M., Linder, J.U., Truffault, V., Schultz, A., et al., (2006). The HAMP domain structure implies helix rotation in transmembrane signaling. *Cell*, **126**, 929–940.
155. Boyken, S.E., Benhaim, M.A., Busch, F., Jia, M.X., Bick, M.J., Choi, H., et al., (2019). De novo design of tunable, pH-driven conformational changes. *Science*,. 364:658-+.
156. Lajoie, M.J., Boyken, S.E., Salter, A.I., Bruffey, J., Rajan, A., Langan, R.A., et al., (2020). Designed protein logic to target cells with precise combinations of surface antigens. *Science*,. 369:1637-+.
157. Nastri, F., D'Alonzo, D., Leone, L., Zambrano, G., Pavone, V., Lombardi, A., (2019). Engineering metalloprotein functions in designed and native scaffolds. *Trends Biochem. Sci.*, **44**, 1022–1040.
158. Webster, A.M., Peacock, A.F.A., (2021). De novo designed coiled coils as scaffolds for lanthanides, including novel imaging agents with a twist. *Chem. Commun.*,.
159. Woolfson, D.N., Alber, T., (1995). Predicting oligomerization states of coiled coils. *Protein Sci.*, **4**, 1596–1607.
160. Testa, O.D., Moutevelis, E., Woolfson, D.N., (2009). CC plus : a relational database of coiled-coil structures. *Nucleic Acids Res.*, **37** D315-D22.
161. Richards, F.M., Lim, W.A., (1993). An analysis of packing in the protein-folding problem. *Q. Rev. Biophys.*, **26**, 423–498.
162. Woolfson, D.N., (2001). Core-directed protein design. *Curr. Opin. Struc. Biol.*, **11**, 464–471.
163. Desjarlais, J.R., Handel, T.M., (1995). New strategies in protein design. *Curr. Opin. Biotech.*, **6**, 460–466.
164. Dahiyat, B.I., Sarisky, C.A., Mayo, S.L., (1997). De novo protein design: Towards fully automated sequence selection. *J. Mol. Biol.*, **273**, 789–796.

165. Dahiyat, B.I., Mayo, S.L., (1996). Protein design automation. *Protein Sci.*, **5**, 895–903.
166. Krivov, G.G., Shapovalov, M.V., Dunbrack, R.L., (2009). Improved prediction of protein side-chain conformations with SCWRL4. *Proteins*, **77**, 778–795.
167. Kajava, A.V., (2012). Tandem repeats in proteins: From sequence to structure. *J. Struct. Biol.*, **179**, 279–288.
168. Offer, G., Hicks, M.R., Woolfson, D.N., (2002). Generalized crick equations for modeling noncanonical coiled coils. *J. Struct. Biol.*, **137**, 41–53.
169. Doyle, L., Hallinan, J., Bolduc, J., Parmeggiani, F., Baker, D., Stoddard, B.L., et al., (2015). Rational design of alpha-helical tandem repeat proteins with closed architectures. *Nature*,. 528:585-+.
170. Brunette, T.J., Parmeggiani, F., Huang, P.S., Bhabha, G., Ekiert, D.C., Tsutakawa, S.E., et al., (2015). Exploring the repeat protein universe through computational protein design. *Nature*,. 528:580-+.
171. Fallas, J.A., Ueda, G., Sheffler, W., Nguyen, V., McNamara, D.E., Sankaran, B., et al., (2017). Computational design of self-assembling cyclic protein homo-oligomers. *Nat. Chem.*, **9**, 353–360.
172. Huang, P.S., Feldmeier, K., Parmeggiani, F., Velasco, D. A.F., Hocker, B., Baker, D., (2016). De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat. Chem. Biol.*,. 12:29-+..
173. Abbass, J., Nebel, J.C., (2020). Rosetta and the journey to predict proteins' structures, 20 years on. *Curr. Bioinform.*, **15**, 611–628.
174. Huang, P.S., Ban, Y.E.A., Richter, F., Andre, I., Vernon, R., Schief, W.R., et al., (2011). RosettaRemodel: A generalized framework for flexible backbone protein design. *PLoS ONE*, **6**
175. Mackenzie, C.O., Zhou, J.F., Grigoryan, G., (2016). Tertiary alphabet for the observable protein structural universe. *P. Natl. Acad. Sci. U.S.A.*, **113** E7438-E47.
176. Frappier, V., Jenson, J.M., Zhou, J.F., Grigoryan, G., Keating, A.E., (2019). Tertiary structural motif sequence statistics enable facile prediction and design of peptides that bind anti-apoptotic Bfl-1 and Mcl-1. *Structure*,. 27:606-+.
177. Jacobs, T.M., Williams, B., Williams, T., Xu, X., Eletsky, A., Federizon, J.F., et al., (2016). Design of structurally distinct proteins using strategies inspired by evolution. *Science*, **352**, 687–690.
178. Pan, X.J., Thompson, M.C., Zhang, Y., Liu, L., Fraser, J. S., Kelly, M.J.S., et al., (2020). Expanding the space of protein geometries by computational design of de novo fold families. *Science*,. 369:1132-+..
179. Koga, N., Tatsumi-Koga, R., Liu, G.H., Xiao, R., Acton, T. B., Montelione, G.T., et al., (2012). Principles for designing ideal protein structures. *Nature*, **491**, 222–227.
180. Voet, A.R.D., Noguchi, H., Addy, C., Simoncini, D., Terada, D., Unzai, S., et al., (2014). Computational design of a self-assembling symmetrical beta-propeller protein. *P. Natl. Acad. Sci. U.S.A.*, **111**, 15102–15107.
181. Marcos, E., Basanta, B., Chidyausiku, T.M., Tang, Y.F., Oberdorfer, G., Liu, G.H., et al., (2017). Principles for designing proteins with cavities formed by curved beta sheets. *Science*, **355**, 201–206.
182. Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., et al., (2010). Predicting protein structures with a multiplayer online game. *Nature*, **466**, 756–760.
183. Koepnick, B., Flatten, J., Husain, T., Ford, A., Silva, D.A., Bick, M.J., et al., (2019). De novo protein design by citizen scientists. *Nature*,. 570:390-+..
184. Sieber, V., Pluckthun, A., Schmid, F.X., (1998). Selecting proteins with improved stability by a phage-based method. *Nat. Biotechnol.*, **16**, 955–960.
185. Kristensen, P., Winter, G., (1998). Proteolytic selection for protein folding using filamentous bacteriophages. *Fold Des.*, **3**, 321–328.
186. Finucane, M.D., Woolfson, D.N., (1999). Core-directed protein design. II. Rescue of a multiply mutated and destabilized variant of ubiquitin. *Biochemistry-Us.*, **38**, 11613–11623.
187. Finucane, M.D., Tuna, M., Lees, J.H., Woolfson, D.N., (1999). Core-directed protein design. I. An experimental method for selecting stable proteins from combinatorial libraries. *Biochemistry-Us.*, **38**, 11604–11612.
188. Polizzi, N.F., Wu, Y.B., Lemmin, T., Maxwell, A.M., Zhang, S.Q., Rawson, J., et al., (2017). De novo design of a hyperstable non-natural protein-ligand complex with sub-angstrom accuracy. *Nat. Chem.*, **9**, 1157–1164.
189. Polizzi, N.F., DeGrado, W.F., (2020). A defined structural unit enables de novo design of small-molecule-binding proteins. *Science*,. 369:1227-+.
190. Watkins, D.W., Jenkins, J.M.X., Grayson, K.J., Wood, N., Steventon, J.W., Le Vay, K.K., et al., (2017). Construction and in vivo assembly of a catalytically proficient and hyperthermostable de novo enzyme. *Nat. Commun.*,. 8: ARTN 358.
191. Lee, M., Wang, T., Makhlynets, O.V., Wu, Y.B., Polizzi, N. F., Wu, H.F., et al., (2017). Zinc-binding structure of a catalytic amyloid from solid-state NMR. *P. Natl. Acad. Sci. U.S.A.*, **114**, 6191–6196.
192. Dumorne, K., Cordova, D.C., Astorga-Elo, M., Renganathan, P., (2017). Extremozymes: A potential source for industrial applications. *J. Microbiol. Biotechn.*, **27**, 649–659.
193. Signarvic, R.S., DeGrado, W.F., (2003). De novo design of a molecular switch: Phosphorylation-dependent association of designed peptides. *J. Mol. Biol.*, **334**, 1–12.
194. Pandya, M.J., Cerasoli, E., Joseph, A., Stoneman, R.G., Waite, E., Woolfson, D.N., (2004). Sequence and structural duality: Designing peptides to adopt two stable conformations. *J. Am. Chem. Soc.*, **126**, 17016–17024.
195. Cerasoli, E., Sharpe, B.K., Woolfson, D.N., (2005). ZiCo: A peptide designed to switch folded state upon binding zinc. *J. Am. Chem. Soc.*, **127**, 15008–15009.
196. Ambroggio, X.I., Kuhlman, B., (2006). Computational design of a single amino acid sequence that can switch between two distinct protein folds. *J. Am. Chem. Soc.*, **128**, 1154–1161.
197. Ambroggio, X.I., Kuhlman, B., (2006). Design of protein conformational switches. *Curr. Opin. Struc. Biol.*, **16**, 525–530.
198. Harrington, L., Fletcher, J.M., Heermann, T., Woolfson, D. N., Schwille, P., (2021). De novo design of a reversible phosphorylation-dependent switch for membrane targeting. *Nat. Commun.*, **12**, 1472.
199. Davey, J.A., Damry, A.M., Goto, N.K., Chica, R.A., (2017). Rational design of proteins that exchange on functional timescales. *Nat. Chem. Biol.*,. 13:1280-+.
200. Chen, K.Y.M., Keri, D., Barth, P., (2020). Computational design of G protein-coupled receptor allosteric signal transductions. *Nat. Chem. Biol.*,. 16:77-+.

201. Vucinic, J., Simoncini, D., Ruffini, M., Barbe, S., Schiex, T., (2020). Positive multistate protein design. *Bioinformatics*, **36**, 122–130.
202. Ferguson, A.L., Ranganathan, R., (2021). 100th anniversary of macromolecular science viewpoint: data-driven protein design. *ACS Macro Lett.*, **10**, 327–340.
203. Baker, D., (2019). What has de novo protein design taught us about protein folding and biophysics?. *Protein Sci.*, **28**, 678–683.
204. Lalaurie, C.J., Dufour, V., Meletiou, A., Ratcliffe, S., Harland, A., Wilson, O., et al., (2018). The de novo design of a biocompatible and functional integral membrane protein using minimal sequence complexity. *Sci. Rep.-Uk*, **8**
205. Curnow, P., Hardy, B.J., Dufour, V., Arthur, C.J., Stenner, R., Hodgson, L.R., et al., (2020). Small-residue packing motifs modulate the structure and function of a minimal de novo membrane protein. *Sci. Rep.-Uk*, **10**
206. Niitsu, A., Heal, J.W., Fauland, K., Thomson, A.R., Woolfson, D.N., (2017). Membrane-spanning alpha-helical barrels as tractable protein-design targets. *Philos. T. R. Soc. B*, **372**
207. Liu, J., Zheng, Q., Deng, Y.Q., Cheng, C.S., Kallenbach, N.R., Lu, M., (2006). A seven-helix coiled coil. *P. Natl. Acad. Sci. U.S.A.*, **103**, 15457–15462.
208. Glasgow, A.A., Huang, Y.M., Mandell, D.J., Thompson, M., Ritterson, R., Loshbaugh, A.L., et al., (2019). Computational design of a modular protein sense-response system. *Science*,. 366:1024-+.
209. Herud-Sikimic, O., Stiel, A.C., Kolb, M., Shanmugaratnam, S., Berendzen, K.W., Feldhaus, C., et al., (2021). A biosensor for the direct visualization of auxin. *Nature*,. 592:768-+.
210. Klima, J.C., Doyle, L.A., Lee, J.D., Rappleye, M., Gagnon, L.A., Lee, M.Y., et al., (2021). Incorporation of sensing modalities into de novo designed fluorescence-activating proteins. *Nat. Commun.*,. 12:ARTN 856.