ACCOUNTS of chemical research

Article

pubs.acs.org/accounts

Miniprotein Design: Past, Present, and Prospects

Published as part of the Accounts of Chemical Research special issue "Chemical Biology of Peptides".

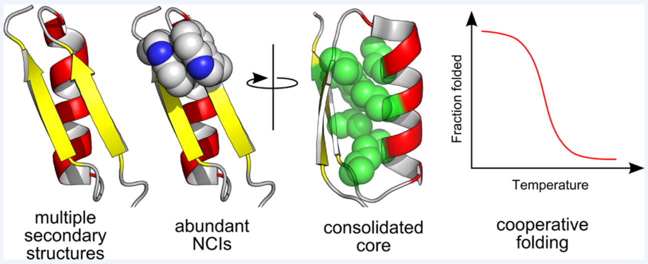
Emily G. Baker,[†] Gail J. Bartlett,[†] Kathryn L. Porter Goff,[†] and Derek N. Woolfson*[,†,‡,§]

[†]School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, U.K.

[‡]School of Biochemistry, University of Bristol, Biomedical Sciences Building, University Walk, Bristol BS8 1TD, U.K.

[§]BrisSynBio and the Bristol BioDesign Institute, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, U.K.

CONSPECTUS: The design and study of miniproteins, that is, polypeptide chains <40 amino acids in length that adopt defined and stable 3D structures, is resurgent. Miniproteins offer possibilities for reducing the complexity of larger proteins and so present new routes to studying sequence-to-structure and sequence-to-stability relationships in proteins generally. They also provide modules for *protein design by pieces* and, with this, prospects for building more-complex or even entirely new protein structures. In addition, miniproteins are useful scaffolds for templating functional domains, for example, those involved in protein–protein interactions, catalysis, and biomolecular binding, leading to potential applications in biotechnology and medicine.



Here we select examples from almost four decades of miniprotein design, development, and dissection. Simply because of the word limit for this Account, we focus on miniproteins that are cooperatively folded monomers in solution and not stabilized by cross-linking or metal binding. In these cases, the optimization of noncovalent interactions is even more critical for the maintenance of the folded states than in larger proteins. Our chronology and catalogue highlights themes in miniproteins, which we explore further and begin to put on a firmer footing through an analysis of the miniprotein structures that have been deposited in the Protein Data Bank (PDB) thus far.

Specifically, and compared with larger proteins, miniproteins generally have a lower proportion of residues in regular secondary structure elements (α helices, β strands, and polyproline-II helices) and, concomitantly, more residues in well-structured loops. This allows distortions of the backbone enabling mini-hydrophobic cores to be made. This also contrasts with larger proteins, which can achieve hydrophobic cores through tertiary contacts between distant regions of sequence. On average, miniproteins have a higher proportion of aromatic residues than larger proteins, and specifically electron-rich Trp and Tyr, which are often found in combination with Pro and Arg to render networks of CH$-\pi$ or cation$-\pi$ interactions. Miniproteins also have a higher proportion of the long-chain charged amino acids (Arg, Glu, and Lys), which presumably reflects salt-bridge formation and their greater surface area-to-volume ratio. Together, these amino-acid preferences appear to support greater densities of noncovalent interactions in miniproteins compared with larger proteins.

We anticipate that with recent developments such as parametric protein design, it will become increasingly routine to use computation to generate and evaluate models for miniproteins *in silico* ahead of experimental studies. This could include accessing new structures comprising secondary structure elements linked in previously unseen configurations. The improved understanding of the noncovalent interactions that stabilize the folded states of such miniproteins that we are witnessing through both in-depth bioinformatics analyses and experimental testing will feed these computational protein designs. With this in mind, we can expect a new and exciting era for miniprotein design, study, and application.

■ INTRODUCTION

Polypeptide chains fold into myriad three-dimensional shapes determined by their amino acid sequences. Currently, there are approaching 120 000 protein structures in the Protein Data Bank (PDB), the chains of which adopt over 1300 distinct folds and have an average length of 800 residues. Understanding this process is known as the protein-folding problem,[1] which has three aspects: How do proteins fold mechanistically? How do proteins fold *in vivo*? And, what information in the amino-acid sequence encodes 3D structure and function?[2] These are complex problems because the folded states are determined by the interplay of many weak noncovalent interactions between thousands of atoms.[3] Moreover, the entropic cost of folding protein chains is only just outweighed by the enthalpy of forming these interactions making folded proteins only marginally stable. One way to address these various aspects of the protein-folding problem is to study so-called miniproteins where complexity is reduced, providing accessible

Received: April 14, 2017
Published: August 23, 2017

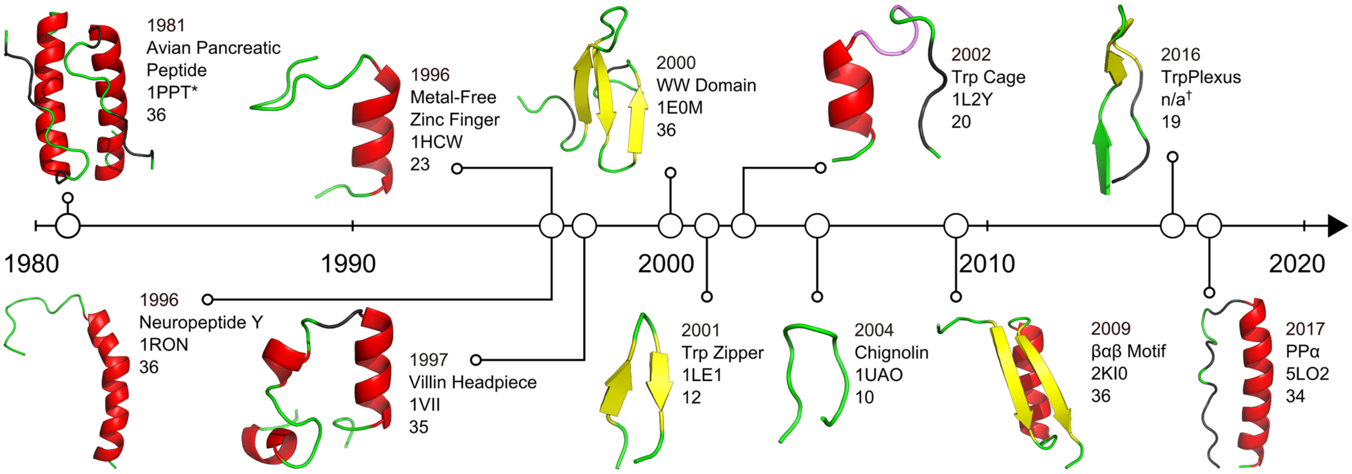


Figure 1. Timeline of miniprotein discovery. Structures are colored by secondary structure: red, α helix; yellow, β strand; dark gray, polyproline-II helix; violet, 3_{10} helix; and green, loop. Annotations: year published; structure type; PDB code; and number of residues. Structures are oriented N → C terminus left to right. * = X-ray crystal structure; all others are solution NMR structures. †See ref 22.

platforms for dissecting contributions to protein folding and stability both *in silico* and *in vitro*.[4,5]

Here we define miniproteins as short proteins of ≤40 amino acids with well-defined folds consisting of two or more secondary structure elements, sequestered hydrophobic cores, and cooperative folding. Standalone, small and cooperatively folded secondary structures exist, for example, single α-helical peptides;[6] however these lack the hydrophobic cores typical of globular proteins. This definition resonates with the concept of foldamers more generally.[7]

The small size of miniproteins necessarily means smaller hydrophobic cores and fewer noncovalent interactions than found in proteins generally. Therefore, many miniproteins are stabilized by metal binding or covalent cross-linking; for example, EF hands,[8] zinc fingers,[9] and cysteine-knot peptides.[10] Here, we focus on water-soluble and largely monomeric miniproteins stabilized solely by noncovalent interactions and without bound metals or covalent cross-links, although we recognize that great strides have been and continue to be made with these miniproteins more generally.[11,12] We give a brief history of the subfield, highlight key examples and themes in more detail, and tease out rules for miniprotein design supported by an analysis of the PDB. Finally, we discuss potential applications and outlook for miniproteins.

Many of the miniproteins described over the last four decades are fragments of larger globular proteins and have been subject to iterative redesign and optimization to enhance stability and impart function.[5,13−15] Successes in this have taught us some general rules of thumb for miniprotein design that we discuss with examples here. However, and despite this, the field of miniprotein design is far from mature.

Most recently, high-throughput methods have been applied to miniprotein design.[16,17] This starts with fragment-based computational design of backbones followed by the generation of libraries of best-fitting sequences. Experimentally, the miniprotein libraries are displayed on yeast, and unstable and stable variants are distinguished by protease treatment and then identified by fluorescence-activated cell sorting and deep sequencing. This reveals sequence-to-stability relationships, which, reassuringly, mirror some well-established rules of protein folding.

A CHRONOLOGY AND COLLECTION OF MINIPROTEINS

Pancreatic Polypeptides

The polyproline-II−loop−α-helix fold was first observed in the X-ray crystal structure of avian pancreatic peptide hormone (aPP) dimer.[18] This compact fold is stabilized by the interdigitation of proline residues from the polyproline-II helix with aromatic residues presented by the α helix to form a hydrophobic core; in addition, π-stacking interactions stabilize the dimer interface, Figure 1. Recently, we have designed a monomeric 34-residue miniprotein with the same overall topology of the pancreatic polypeptides,[19] which we call PPα and discuss later, Figures 2A and 3.

Directed evolution of synthetic aPP monomers has been used to develop miniprotein-based ligands as therapeutics. Optimization of the polyproline-II helix in aPP gave a variant with high affinity for the ActA target protein in *Listeria monocytogenes*, EVH1 mena$_{1-112}$. Importantly, the miniprotein discriminates between paralogs and reduces bacterial motility.[13] A similar strategy has been used to optimize the α-helix of aPP for sequence-specific DNA recognition,[14] and the introduction of arginine residues in aPP facilitates transport of the miniprotein into cells.[20] Artificial esterases have also been developed by grafting catalytic residues onto the solvent-exposed α helix of bovine pancreatic polypeptide, bPP.[21]

$\beta\beta\alpha$ Folds and Metal-Free Zinc Fingers

Small, independent $\beta\beta\alpha$ units are best exemplified by DNA-binding zinc fingers. These were first identified in transcription factor IIIA from *Xenopus* oocytes, and an NMR structure determined for the Xfin domain followed.[23] The fold, which comprises a β hairpin, a connecting loop, and an α helix, is not driven by the conserved hydrophobic core but by the binding of zinc usually through His_2Cys_2 motifs. The development and modular assembly of metal-binding zinc-finger domains has created artificial proteins and enzymes that can recognize defined regions of DNA for the activation, repression, or alteration of user-specified genes and has contributed early in the development of genome editing.[24]

Regarding metal-free designs, a 23-residue monomeric structure has been achieved through iterative design enhancing a hydrophobic core, α-helix structure, and inclusion of a

DOI: 10.1021/acs.accounts.7b00186
Acc. Chem. Res. 2017, 50, 2085−2092

suitable turn.[25] A computational design has produced a weakly cooperatively folded peptide with a midpoint of unfolding (T_M) of 39 °C.[26] The broad transition is consistent with a low enthalpy of folding expected for a small hydrophobic core.

Villin Headpiece

Another approach to miniprotein design is to pare down larger natural proteins. The Villin headpiece, a 35-residue fragment of the chicken protein (HP-35), folds in water.[27] The NMR ensemble reveals three α-helical segments with each helix contributing residues to a central hydrophobic core.[27] HP-35 is surprisingly thermostable with a T_M of 70 °C.[28]

β-Hairpins and Trp-Zippers

Early examples of free-standing β-structures based on natural fragments were only moderately folded in fully aqueous media.[32,33] However, designed Trp-zippers are well-folded examples of short (12−16 residues) *de novo* β hairpins with interlocked tryptophans and right-handed highly twisted strands, Figure 2D.[29] They have exceptional thermal stabilities and reversible cooperative unfolding.

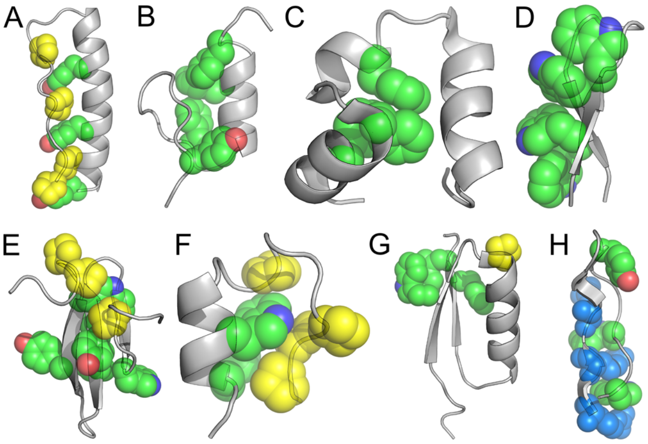


Figure 2. Miniprotein structures. (A) PPα-Tyr (PDB: 5LO2).[19] (B) $\beta\beta\alpha$ metal-free zinc finger (PDB: 1PSV).[26] (C) Villin headpiece (PDB: 1YRI).[28] (D) Trp-zipper β-hairpin (PDB: 1LE3).[29] (E) Protype WW domain (PDB: 1E0M).[30] (F) Trp-cage (PDB: 2J0F).[31] (G) $\beta\alpha\beta$ (PDB: 2KI0). (H) TrpPlexus.[22] Structures are oriented N → C terminus left to right. Key: aromatics Trp, Tyr, and Phe, green; Pro, yellow; and Arg, light blue.

Trp-pocket β hairpins are stabilized via cation−π interactions in which a single Lys packs against a diTrp cleft on the opposite strand.[34] Some 12-residue Trp-pockets are fully folded[34] and resist degradation,[35] making them the most stable β hairpins reported.

From a wealth of studies, reliable guidelines for the design, optimization, and stabilization of monomeric β-hairpins include a hydrophobic cluster on one surface of the hairpin and close to the loop;[36] interstrand side-chain interactions, particularly Trp−Trp;[29,37] high turn and β-sheet propensities;[38] and charged or aromatic residues or β-capping motifs to secure the termini via cross-strand interactions.[39]

Designed Three-Stranded β-Sheets

A challenge in the design of isolated β-sheets is to avoid β-amyloid-like assemblies.[40] Some tentative three-stranded antiparallel β-sheets have been achieved in aqueous methanol by appending strands onto β-hairpins.[41] An early NMR structure of "Betanova" in water incorporates an aromatic-rich hydrophobic cluster on one surface of the sheet.[42] However, subsequent studies indicate only partial folding, though improved stability has been achieved by computationally informed mutations.[43]

Protein redesign presents another route to β-sheet miniproteins. WW domains are natural antiparallel 3-stranded β sheets, named after the two conserved Trp residues in the first and third β strands.[44] X-ray crystal structures of two shorter (34 and 37 residues) natural WW domains followed, in addition to that for a designed 33-residue prototype, Figure 2E.[30] Natural WW domains have been engineered for alternative functions, for example, the incorporation of a DNA binding pocket,[15] and for probing carbohydrate−aromatic packing interactions.[5]

Trp-Cage

The Trp-cage is a 20-residue miniprotein from the gila monster extendin-4.[45] The original truncation is only folded in aqueous trifluoroethanol.[45] However, NMR structures of variants show a well-ordered fold comprising an α helix followed by a well-structured loop, with a hydrophobic core centered on a single Trp residue buttressed by Pro side chains, Figure 2F.[31] Owing to its small size and wealth of experimental data now available, the Trp-cage has become a paradigm for experimental and computational miniprotein folding.[46]

$\beta\alpha\beta$ Designs

$(\beta\alpha)_n$ repeats occur widely in natural proteins, for example, TIM barrels, the Rossmann fold, and Leucine Rich Repeats. The alternating secondary structure elements lead to parallel β-sheets. The first *de novo* designed, standalone, water-soluble $\beta\alpha\beta$ unit comprises a 12-residue α helix paired with two 5-residue β strands via a small hydrophobic core.[47] Although initial designs were molten globule, a folded state has been stabilized by installing a Trp-zip-like Trp pair in the β-strands.[29] An NMR structure of the resulting 36-residue $\beta\alpha\beta$ construct (Figure 2G) confirms face-to-face packing of the installed pair. The miniprotein is highly stable up to temperatures of 90 °C, which is remarkable for a small miniprotein with only proteinogenic amino acids and without covalent cross-links.

TrpPlexus

The first miniprotein comprising a β-strand and a polyproline-II helix has been achieved through a fragment-based design, Figure 2H.[22] TrpPlexus combines a short, Arg-rich, N-terminal β strand, and a C-terminal polyproline-II helix that is free of Pro but rich in Trp in a WSXWS motif. These are borrowed from a fibronectin III binding domain, and linked by a D-Pro-Gly loop to give a 19-residue construct. NMR spectroscopy shows the Arg and Trp residues interdigitate to form a cation−π network.[22] A disulfide-cyclized TrpPlexus tolerates N-substituted Gly and Pro residues in the polyproline-II helix, opening up potential for proteinogenic and peptoid side-chain placements for peptidomimetic inhibitors of protein−protein interactions.[48]

PPα

Recently we combined fragment-based and rational design to create the monomeric miniprotein PPα.[19] PPα has a polyproline-II helix−loop−α-helix topology that is stabilized by interdigitation of Pro residues of the polyproline helix into aromatic residues presented by the α-helix, similar to knobs-into-holes interactions found in coiled coils, Figure 3. The helices were borrowed from the bacterial adhesin AgI/II[49] and an intervening loop from a pancreatic polypeptide.[50] We

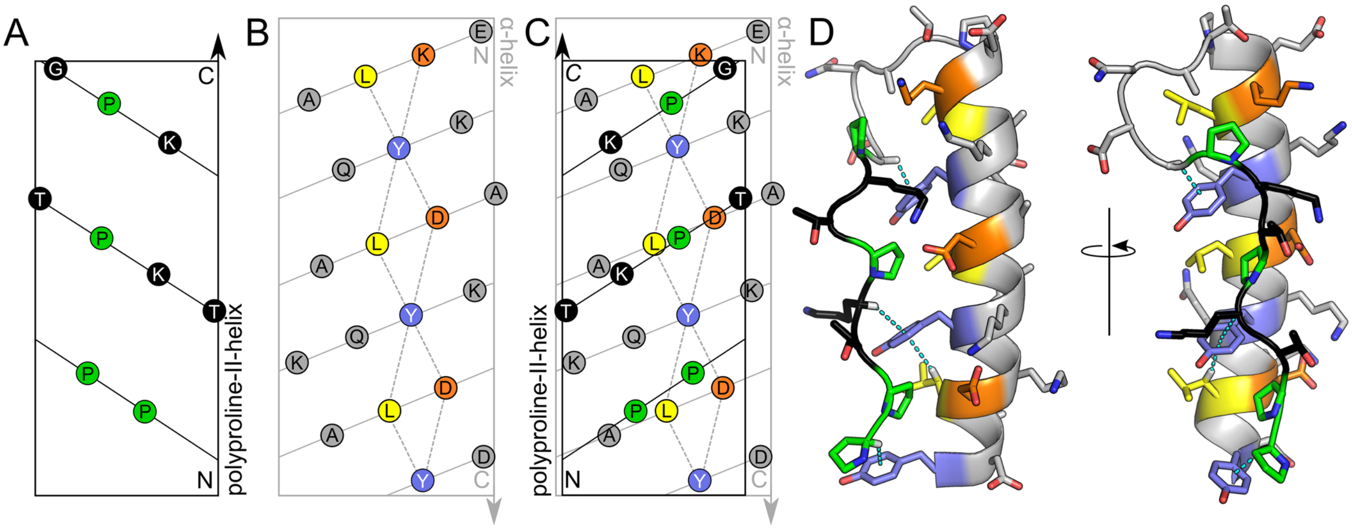


Figure 3. Design of PPα combining polyproline-II and α-helices. (A, B) 2D helical net representations (i.e., Cα atoms mapped onto cylinder of appropriate radii) of a canonical polyproline-II helix (A) and α-helix (B) and (C) these two nets overlaid showing "knobs-into-holes" packing of Pro and Tyr side chains. (D) Representative NMR structure of PPα-Tyr showing the CH−π interactions found (PDB: 5LO2).[19] Figure adapted from ref 19.

selected ∼6 turns of α helix to partner ∼3 turns of polyproline helix to maintain knobs-into-holes-like interactions along the lengths of both helices. The tyrosine variant, PPα-Tyr, is water-soluble and monomeric and unfolds cooperatively with a T_M of 39 °C.

An NMR structure reveals intimate CH−π interactions[51] between the proline and aromatic side chains, Figure 3D. We have explored the importance of these interactions by mutating the three parent tyrosine residues to *para*-substituted phenylalanine residues with varying ring electron densities. These experiments highlight electronic and electrostatic contributions to the interaction beyond van der Waals' contacts, as the more electron-rich aromatics lead to more stable PPα variants. Interestingly, of the proteinogenic aromatic side chains, Pro−Tyr and Pro−Trp interactions gave PPα variants with similar thermal stabilities. This corroborates our bioinformatics analyses of the PDB, which highlights a preference for CH−π interactions between Pro and the electron-rich aromatics but not Phe.[19]

■ COMMON FEATURES: BIOINFORMATICS ANALYSIS OF MINIPROTEINS IN THE PDB

The above examples hint at common features that relate miniprotein sequence, structure, and stability. For example, the importance of large aromatics, particularly Trp, in the hydrophobic cores of miniproteins. To explore this and to seek other sequence-to-structure relationships, we performed a comparative bioinformatics analysis of the mini- and large proteins in the PDB. For this, we culled a nonredundant database of X-ray crystal and solution NMR structures of miniproteins of ≤40 residues and with <40% pairwise sequence identity, Figure 4 and a set of larger proteins with >100 residues and high-resolution (≤1.0 Å) X-ray crystal structures. We verified our set of miniproteins contained only monomeric structures determined in aqueous media.

First, we compared the amino-acid compositions of the two sets, Figure 5. Miniproteins use three classes of amino acids more often than the larger proteins: long or charged amino acids, for example, Arg, Glu, Lys, and Met (and to a lesser extent His); electron-rich aromatics, Trp and Tyr, with Trp showing a preference for miniproteins twice that of larger proteins; and Pro. In contrast, small amino acids and the aliphatic hydrophobics are found more often in the larger proteins. This suggests that the polar aromatics and longer charged amino acids provide routes to good noncovalent interactions, that is, salt bridges and CH−π and cation−π interactions, and that Pro helps reduce the entropic cost of folding as well as buttressing aromatics in the core.

We have also examined secondary structure content using the Define Secondary Structure of Proteins (DSSP) algorithm, Figure 4.[52] This revealed similar proportions of α helix in miniproteins and larger proteins; but less β strand in the former (with the exception of the water-soluble β-hairpins and three-stranded β-sheets, of course). It is possible that this is due to bias in the data set rather than miniprotein requirements. Also, it appears that miniproteins have more-contorted backbones and make better use of structured loops to best sequester hydrophobics within their interiors. Concomitant with reduced regular secondary structures, we found fewer main chain−main chain hydrogen bonds in miniproteins, with half of the residues making these compared with approximately three-quarters in larger proteins. The shortfall in miniproteins is likely made up by more hydrogen bonds to water, as their small size gives greater surface area-to-volume ratios.

So how are miniproteins stabilized? True, the overall entropic cost of folding a miniprotein might be lower than that for larger proteins, but it will still be unfavorable and will have to be recouped by enthalpically favorable interactions. However, it is clear that fewer such interactions are made in miniproteins as they generally have broader thermal unfolding transitions than larger proteins.[19] Nonetheless, there must be favorable noncovalent interactions to outweigh the $T\Delta S$ term of folding. To address this, we are actively interrogating the above databases for sequence-to-structure/stability relationships. It is early in this analysis, but trends are emerging. For example, we find that, when normalized for length, miniproteins make up to eight times as many salt bridges as their longer counterparts; and when normalized for both length and

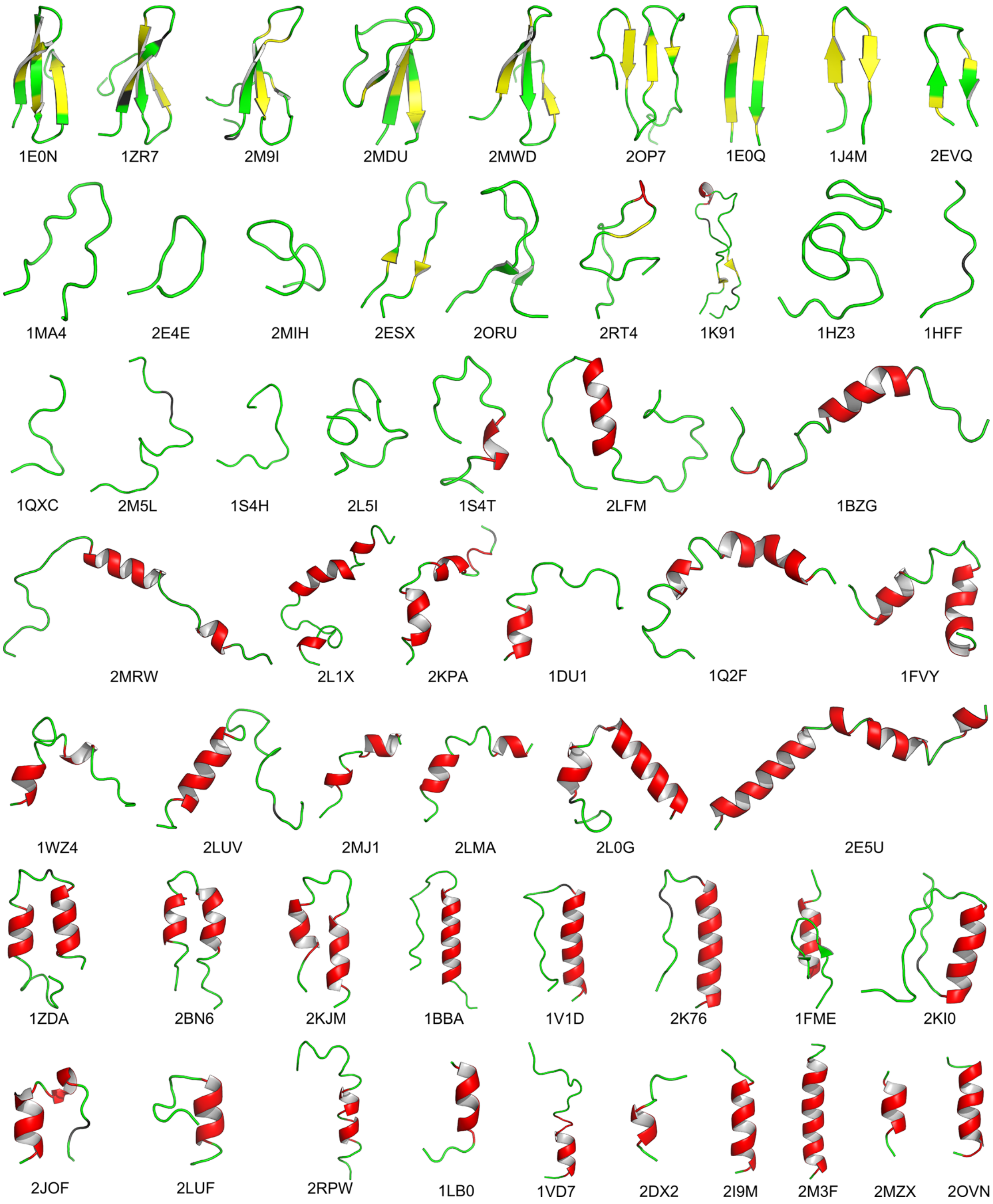


Figure 4. Nonredundant miniproteins. X-ray crystal and solution NMR structures of miniproteins downloaded from the PDB on 23/02/2017. Structures are arranged loosely by conformation, and oriented N → C terminus left to right. For the solution-phase structures, this set includes only those determined in purely aqueous media. Key: α helix, red; β strand, yellow; loop, green; and polyproline-II helix, gray. Note: DSSP assigns some regions of β hairpins and three-stranded β sheets as unstructured (green).

number of aromatic residues, miniproteins are approximately six-times denser in CH−π interactions.

■ OUTLOOK

We have attempted to cover four decades of miniprotein research in a 6000-word Account. As a result, we have had to omit many of the fascinating aspects of these studies, including details of the approaches taken, the experimental and theoretical methods used, the sequences explored, and the nuances of the results obtained. However, we hope to have conveyed two main aspects of miniprotein research and development that others may find useful, namely: first, a chronology and overview of miniprotein discovery and, with it, the design and engineering approaches taken to reveal them; and, second, common sequence and structural features found in miniproteins through these studies and our own, albeit preliminary, bioinformatics analysis of the RCSB PDB.

On the latter, sequence-to-structure/stability relationships are clearly emerging for miniprotein folding. Moreover, these are being discussed and, indeed, understood in terms of the noncovalent interactions that underpin them. For miniproteins, and mostly in contrast to what is possible with larger proteins,

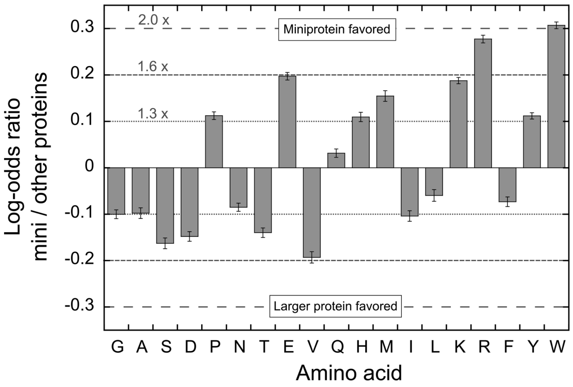


Figure 5. Amino-acid usage in miniproteins vs larger proteins. These were calculated as the $\log_{10}$ ratio of the normalized propensity of amino acids in miniproteins (n = 55) vs proteins >100 residues long (n = 120) Ratios of ±0.3 indicate a 2-fold preference (long dashed lines); ±0.2 a 1.6-fold preference (short dashed lines); and ±0.1 a 1.3-fold preference (dotted lines). Amino acids are ordered by increasing side chain volume.

these interactions can be probed with atomic resolution using synthetic peptide chemistry: nonproteinogenic amino acids can be introduced into miniproteins, and principles and methods from physical organic chemistry can be used to rationalize their impact on structure and stability.[5,19] Such knowledge and understanding will not only illuminate how miniproteins are stabilized but also how protein structures are specified and maintained in general.[5,19,53,54] In turn, this will undoubtedly improve our abilities to engineer existing miniproteins and to design new examples of these *de novo*. We anticipate that, along with the *in biro* (back-of-the-envelope) and rational design approaches that have been favored to date, parametric computational protein design will become increasingly used to deliver new miniproteins. Here, a challenge for the protein-design community will be to make their methods more available to nonexpert users.

One thing that we have neglected in this Account is the potential to functionalize miniproteins for both basic and applied research; we have had to focus on structures and sequence-to-structure/stability relationships rather than on structure−function relationships. Many have contributed to this important aspect of the field.[5,13−15] For example, miniproteins provide scaffolds onto which functional motifs can be grafted.[14,55] The most clear-cut application here is the introduction of binding and recognition motifs, for example, to interfere with protein−protein interactions. In addition, and although more challenging, there are prospects for introducing catalytic functions into simplified peptide and protein architectures.[14,21] With the array of structures that we have presented and others that are coming online, and with the underpinning thermodynamic understanding of these, we anticipate further and considerable advances on this road to functional miniproteins.

Finally, and particularly exciting to us, is the concept that miniproteins might be used as building blocks to design and engineer entirely new protein folds. This might be termed *proteins from peptides* (we thank Andrei Lupas for this phrase) or *protein design by pieces*.[56−58] It offers routes into what is being called the *dark matter of protein fold space*,[59] and, thus, into a truly synthetic biology.[60]

AUTHOR INFORMATION

Corresponding Author

*D.N.W. E-mail: D.N.Woolfson@bristol.ac.uk).

ORCID

Derek N. Woolfson: 0000-0002-0394-3202

Author Contributions

All authors contributed equally to this Account.

Notes

The authors declare no competing financial interest.
Lists of PDB codes used for analysis along with raw data presented in Figure 5 are available to download (http://dx.doi.org/10.5523/bris.1gefaj9sba0dt2fhglr6fyicjh).

Biographies

Emily G. Baker received a Chemistry degree and then her Ph.D. from the University of Bristol. Her Ph.D. was on the design of single α-helical peptides and the understanding of electrostatic interactions within these. Emily is now a postdoctoral research associate at Bristol working on the *de novo* design of unexplored protein folds with a focus on probing noncovalent interactions within these and using DNA-guided peptide assembly.

Gail J. Bartlett obtained her degree in Biochemistry from the University of Oxford, and her Ph.D. in structural bioinformatics from the University of London. Her current research interests focus on the relationships between protein sequence and structure/function and particularly on the role of noncovalent interactions in protein folding, stability, and design.

Kathryn L. Porter Goff received her M.Sci. in Chemistry from the University of Bristol. She is currently pursuing a Ph.D. within the Bristol Chemical Synthesis CDT under the supervision of Prof. Dek Woolfson working on the rational design, synthesis, and characterization of new protein folds.

Derek N. Woolfson took his first degree in Chemistry at the University of Oxford and gained a Ph.D. in Chemistry and Biochemistry at the University of Cambridge. He then did postdoctoral research at University College London and the University of California, Berkeley. After 10 years as Lecturer through to Professor of Biochemistry at the University of Sussex, he moved to the University of Bristol in 2005 to take up a joint chair in Chemistry and Biochemistry. Dek's research is at the interface between chemistry and biology, applying chemical methods and principles to understand biological phenomena. Specifically, his group is interested in the challenge of rational protein design and how this can be applied in synthetic biology and biotechnology. His particular emphasis is on making completely new protein structures from peptide blocks and peptide-based biomaterials for applications in cell biology and medicine. Dek is also co-Director of BrisSynBio, a BBSRC/EPSRC-funded Synthetic Biology Research Centre.

ACKNOWLEDGMENTS

E.G.B. and D.N.W. are supported by a BBSRC/ERASynBio grant (BB/M005615/1); G.J.B. and D.N.W. are supported by the ERC (340764); K.L.P.G. is supported by the EPSRC-funded Bristol Chemical Synthesis Centre for Doctoral Training (EP/G036764/1); and D.N.W. is a Royal Society Wolfson Research Merit Award holder (WM140008).

REFERENCES

(1) Dill, K. A.; MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* **2012**, *338*, 1042−1046.

(2) Travaglini-Allocatelli, C.; Ivarsson, Y.; Jemth, P.; Gianni, S. Folding and Stability of Globular Proteins and Implications for Function. *Curr. Opin. Struct. Biol.* **2009**, *19*, 3−7.

(3) Dill, K. A. Dominant Forces in Protein Folding. *Biochemistry* **1990**, *29*, 7133−7155.

(4) Gellman, S. H.; Woolfson, D. N. Mini-Proteins Trp the Light Fantastic. *Nat. Struct. Biol.* **2002**, *9*, 408−410.

(5) Chen, W. T.; Enck, S.; Price, J. L.; Powers, D. L.; Powers, E. T.; Wong, C.-H.; Dyson, H. J.; Kelly, J. W. Structural and Energetic Basis of Carbohydrate−Aromatic Packing Interactions in Proteins. *J. Am. Chem. Soc.* **2013**, *135*, 9877−9884.

(6) Baker, E. G.; Bartlett, G. J.; Crump, M. P.; Sessions, R. B.; Linden, N.; Faul, C. F. J.; Woolfson, D. N. Local and Macroscopic Electrostatic Interactions in Single α-Helices. *Nat. Chem. Biol.* **2015**, *11*, 221−228.

(7) Gellman, S. H. Foldamers: A manifesto. *Acc. Chem. Res.* **1998**, *31*, 173−180.

(8) Gifford, J. L.; Walsh, M. P.; Vogel, H. J. Structures and Metal-Ion-Binding Properties of the Ca^{2+}-Binding Helix−Loop−Helix EF-Hand Motifs. *Biochem. J.* **2007**, *405*, 199−221.

(9) Wolfe, S. A.; Nekludova, L.; Pabo, C. O. DNA Recognition by Cys_2His_2 Zinc Finger Proteins. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 183−212.

(10) Craik, D. J.; Daly, N. L.; Waine, C. The Cystine Knot Motif in Toxins and Implications for Drug Design. *Toxicon* **2001**, *39*, 43−60.

(11) Bhardwaj, G.; Mulligan, V. K.; Bahl, C. D.; Gilmore, J. M.; Harvey, P. J.; Cheneval, O.; Buchko, G. W.; Pulavarti, S. V. S. R. K.; Kaas, Q.; Eletsky, A.; Huang, P.-S.; Johnsen, W. A.; Greisen, P., Jr.; Rocklin, G. J.; Song, Y.; Linsky, T. W.; Watkins, A.; Rettie, S. A.; Xu, X.; Carter, L. P.; Bonneau, R.; Olson, J. M.; Coutsias, E.; Correnti, C. E.; Szyperski, T.; Craik, D. J.; Baker, D. Accurate *de novo* Design of Hyperstable Constrained Peptides. *Nature* **2016**, *538*, 329−335.

(12) Yu, F. T.; Cangelosi, V. M.; Zastrow, M. L.; Tegoni, M.; Plegaria, J. S.; Tebo, A. G.; Mocny, C. S.; Ruckthong, L.; Qayyum, H.; Pecoraro, V. L. Protein Design: Toward Functional Metalloenzymes. *Chem. Rev.* **2014**, *114*, 3495−3578.

(13) Golemi-Kotra, D.; Mahaffy, R.; Footer, M. J.; Holtzman, J. H.; Pollard, T. D.; Theriot, J. A.; Schepartz, A. High Affinity, Paralog-Specific Recognition of the Mena EVH1 Domain by a Miniature Protein. *J. Am. Chem. Soc.* **2004**, *126*, 4−5.

(14) Yang, L.; Schepartz, A. Relationship between Folding and Function in a Sequence-Specific Miniature DNA-Binding Protein. *Biochemistry* **2005**, *44*, 7469−7478.

(15) Stewart, A. L.; Park, J. H.; Waters, M. L. Redesign of a WW Domain Peptide for Selective Recognition of Single-Stranded DNA. *Biochemistry* **2011**, *50*, 2575−2584.

(16) Rocklin, G. J.; Chidyausiku, T. M.; Goreshnik, I.; Ford, A.; Houliston, S.; Lemak, A.; Carter, L.; Ravichandran, R.; Mulligan, V. K.; Chevalier, A.; Arrowsmith, C. H.; Baker, D. Global Analysis of Protein Folding Using Massively Parallel Design, Synthesis and Testing. *Science* **2017**, *357*, 168.

(17) Woolfson, D. N.; Baker, E. G.; Bartlett, G. J. How do Miniproteins Fold? *Science* **2017**, *357*, 133.

(18) Blundell, T. L.; Pitts, J. E.; Tickle, I. J.; Wood, S. P.; Wu, C.-W. X-Ray Analysis (1.4-Å Resolution of Avian Pancreatic-Polypeptide: Small Globular Protein Hormone. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 4175−4179.

(19) Baker, E. G.; Williams, C.; Hudson, K. L.; Bartlett, G. J.; Heal, J. W.; Porter Goff, K. L.; Crump, M. P.; Woolfson, D. N.; Sessions, R. B. Engineering Protein Stability with Atomic Precision in a Monomeric Miniprotein. *Nat. Chem. Biol.* **2017**, *13*, 764−770.

(20) Smith, B. A.; Daniels, D. S.; Coplin, A. E.; Jordan, G. E.; McGregor, L. M.; Schepartz, A. Minimally Cationic Cell-Permeable Miniature Proteins via α-Helical Arginine Display. *J. Am. Chem. Soc.* **2008**, *130*, 2948−2949.

(21) Nicoll, A. J.; Allemann, R. K. Nucleophilic and General Acid Catalysis at Physiological pH by a Designed Miniature Esterase. *Org. Biomol. Chem.* **2004**, *2*, 2175−2180.

(22) Craven, T. W.; Cho, M.-K.; Traaseth, N. J.; Bonneau, R.; Kirshenbaum, K. A Miniature Protein Stabilized by a Cation−π Interaction Network. *J. Am. Chem. Soc.* **2016**, *138*, 1543−1550.

(23) Lee, M. S.; Gippert, G. P.; Soman, K. V.; Case, D. A.; Wright, P. E. Three-Dimensional Solution Structure of a Single Zinc Finger DNA-Binding Domain. *Science* **1989**, *245*, 635−637.

(24) Gersbach, C. A.; Gaj, T.; Barbas, C. F. Synthetic Zinc Finger Proteins: The Advent of Targeted Gene Regulation and Genome Modification Technologies. *Acc. Chem. Res.* **2014**, *47*, 2309−2318.

(25) Struthers, M. D.; Cheng, R. P.; Imperiali, B. Design of a Monomeric 23-Residue Polypeptide with Defined Tertiary Structure. *Science* **1996**, *271*, 342−345.

(26) Dahiyat, B. I.; Mayo, S. L. De Novo Protein Design: Fully Automated Sequence Selection. *Science* **1997**, *278*, 82−87.

(27) McKnight, C. J.; Matsudaira, P. T.; Kim, P. S. NMR Structure of the 35-Residue Villin Headpiece Subdomain. *Nat. Struct. Biol.* **1997**, *4*, 180−184.

(28) Chiu, T. K.; Kubelka, J.; Herbst-Irmer, R.; Eaton, W. A.; Hofrichter, J.; Davies, D. R. High-Resolution X-Ray Crystal Structures of the Villin Headpiece Subdomain, an Ultrafast Folding Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 7517−7522.

(29) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. Tryptophan Zippers: Stable, Monomeric β-Hairpins. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 5578−5583.

(30) Macias, M. J.; Gervais, V.; Civera, C.; Oschkinat, H. Structural Analysis of WW Domains and Design of a WW Prototype. *Nat. Struct. Biol.* **2000**, *7*, 375−379.

(31) Barua, B.; Lin, J. C.; Williams, V. D.; Kummler, P.; Neidigh, J. W.; Andersen, N. H. The Trp-Cage: Optimizing the Stability of a Globular Miniprotein. *Protein Eng., Des. Sel.* **2008**, *21*, 171−185.

(32) Blanco, F. J.; Jiménez, M. A.; Herranz, J.; Rico, M.; Santoro, J.; Nieto, J. L. NMR Evidence of a Short Linear Peptide that Folds into a β-Hairpin in Aqueous Solution. *J. Am. Chem. Soc.* **1993**, *115*, 5887−5888.

(33) Cox, J. P. L.; Evans, P. A.; Packman, L. C.; Williams, D. H.; Woolfson, D. N. Dissecting the Structure of a Partially Folded Protein − Circular Dichroism and Nuclear Magnetic Resonance Studies of Peptides from Ubiquitin. *J. Mol. Biol.* **1993**, *234*, 483−492.

(34) Riemen, A. J.; Waters, M. L. Design of Highly Stabilized β-Hairpin Peptides through Cation−π Interactions of Lysine and *N*-Methyllysine with an Aromatic Pocket. *Biochemistry* **2009**, *48*, 1525−1531.

(35) Cline, L. L.; Waters, M. L. The Structure of Well-Folded β-Hairpin Peptides Promotes Resistance to Peptidase Degradation. *Biopolymers* **2009**, *92*, 502−507.

(36) Espinosa, J. F.; Gellman, S. H. A Designed β-Hairpin Containing a Natural Hydrophobic Cluster. *Angew. Chem., Int. Ed.* **2000**, *39*, 2330−2333.

(37) Santiveri, C. M.; Jiménez, M. A. Tryptophan Residues: Scarce in Proteins but Strong Stabilizers of β-Hairpin Peptides. *Biopolymers* **2010**, *94*, 779−790.

(38) Ramírez-Alvarado, M.; Blanco, F. J.; Serrano, L. *De Novo* Design and Structural Analysis of a Model β-Hairpin Peptide System. *Nat. Struct. Biol.* **1996**, *3*, 604−612.

(39) Kiehna, S. E.; Waters, M. L. Sequence Dependence of β-Hairpin Structure: Comparison of a Salt Bridge and an Aromatic Interaction. *Protein Sci.* **2003**, *12*, 2657−2667.

(40) Doig, A. J. A Three Stranded β-Sheet Peptide in Aqueous Solution Containing *N*-Methyl Amino Acids to Prevent Aggregation. *Chem. Commun.* **1997**, 2153−2154.

(41) Sharman, G. J.; Searle, M. S. Cooperative Interaction Between the Three Strands of a Designed Antiparallel β-sheet. *J. Am. Chem. Soc.* **1998**, *120*, 5291−5300.

(42) Kortemme, T.; Ramírez-Alvarado, M.; Serrano, L. Design of a 20-Amino Acid, Three-Stranded β-Sheet Protein. *Science* **1998**, *281*, 253−256.

(43) López, M.; Lacroix, E.; Ramírez-Alvarado, M.; Serrano, L. Computer-Aided Design of β-Sheet Peptides. *J. Mol. Biol.* **2001**, *312*, 229−246.

(44) Macias, M. J.; Hyvönen, M.; Baraldi, E.; Schultz, J.; Sudol, M.; Saraste, M.; Oschkinat, H. Structure of the WW Domain of a Kinase-Associated Protein Complexed with a Proline-Rich Peptide. *Nature* **1996**, *382*, 646−649.

(45) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Designing a 20-Residue Protein. *Nat. Struct. Biol.* **2002**, *9*, 425−430.

(46) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517−520.

(47) Liang, H.; Chen, H.; Fan, K.; Wei, P.; Guo, X.; Jin, C.; Zeng, C.; Tang, C.; Lai, L. De Novo Design of a $\beta\alpha\beta$ Motif. *Angew. Chem., Int. Ed.* **2009**, *48*, 3301−3303.

(48) Craven, T. W.; Bonneau, R.; Kirshenbaum, K. PPII Helical Peptidomimetics Templated by Cation-π Interactions. *ChemBioChem* **2016**, *17*, 1824−1828.

(49) Larson, M. R.; Rajashankar, K. R.; Patel, M. H.; Robinette, R. A.; Crowley, P. J.; Michalek, S.; Brady, L. J.; Deivanayagam, C. Elongated Fibrillar Structure of a Streptococcal Adhesin Assembled by the High-Affinity Association of α- and PPII-Helices. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 5983−5988.

(50) Blundell, T. L.; Pitts, J. E.; Tickle, I. J.; Wood, S. P.; Wu, C. W. X-Ray Analysis (1. 4-Å Resolution) of Avian Pancreatic Polypeptide: Small Globular Protein Hormone. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 4175−4179.

(51) Hudson, K. L.; Bartlett, G. J.; Diehl, R. C.; Agirre, J.; Gallagher, T.; Kiessling, L. L.; Woolfson, D. N. Carbohydrate-Aromatic Interactions in Proteins. *J. Am. Chem. Soc.* **2015**, *137*, 15152−15160.

(52) Touw, W. G.; Baakman, C.; Black, J.; te Beek, T. A. H.; Krieger, E.; Joosten, R. P.; Vriend, G. A Series of PDB-Related Databanks for Everyday Needs. *Nucleic Acids Res.* **2015**, *43*, D364−D368.

(53) Arnold, U.; Raines, R. T. Replacing a Single Atom Accelerates the Folding of a Protein and Increases its Thermostability. *Org. Biomol. Chem.* **2016**, *14*, 6780−6785.

(54) Zondlo, N. J. Aromatic−Proline Interactions: Electronically Tunable CH/π Interactions. *Acc. Chem. Res.* **2013**, *46*, 1039−1049.

(55) Cobos, E. S.; Pisabarro, M. T.; Vega, M. C.; Lacroix, E.; Serrano, L.; Ruiz-Sanz, J.; Martinez, J. C. A Miniprotein Scaffold Used to Assemble the Polyproline II Binding Epitope Recognized by SH3 Domains. *J. Mol. Biol.* **2004**, *342*, 355−365.

(56) Bharat, T. A. M.; Eisenbeis, S.; Zeth, K.; Höcker, B. A $\beta\alpha$-Barrel Built by the Combination of Fragments from Different Folds. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 9942−9947.

(57) Fletcher, J. M.; Boyle, A. L.; Bruning, M.; Bartlett, G. J.; Vincent, T. L.; Zaccai, N. R.; Armstrong, C. T.; Bromley, E. H. C.; Booth, P. J.; Brady, R. L.; Thomson, A. R.; Woolfson, D. N. A Basis Set of *de Novo* Coiled-Coil Peptide Oligomers for Rational Protein Design and Synthetic Biology. *ACS Synth. Biol.* **2012**, *1*, 240−250.

(58) Zhu, H. B.; Sepulveda, E.; Hartmann, M. D.; Kogenaru, M.; Ursinus, A.; Sulz, E.; Albrecht, R.; Coles, M.; Martin, J.; Lupas, A. N. Origin of a Folded Repeat Protein from an Intrinsically Disordered Ancestor. *eLife* **2016**, *5*, e16761.

(59) Taylor, W. R.; Chelliah, V.; Hollup, S. M.; MacDonald, J. T.; Jonassen, I. Probing the "Dark Matter" of Protein Fold Space. *Structure* **2009**, *17*, 1244−1252.

(60) Woolfson, D. N.; Bartlett, G. J.; Burton, A. J.; Heal, J. W.; Niitsu, A.; Thomson, A. R.; Wood, C. W. De Novo Protein Design: How Do We Expand into the Universe of Possible Protein Structures? *Curr. Opin. Struct. Biol.* **2015**, *33*, 16−26.