The University of Osaka
Institutional Knowledge Archive

Title	Inventing Novel Protein Folds
Author(s)	Koga, Nobuyasu; Tatsumi-Koga, Rie
Citation	Journal of Molecular Biology. 2024, 436(21), p. 168791
Version Type	VoR
URL	https://hdl.handle.net/11094/98433
rights	This article is licensed under a Creative Commons Attribution 4.0 International License.
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

Review Article

Inventing Novel Protein Folds

Nobuyasu Koga [1,2,*] and Rie Tatsumi-Koga [1]

*1 - **Laboratory for Protein Design,** Institute for Protein Research (IPR), Osaka University, Suita, Osaka 565-0871, Japan*
*2 - **Protein Design Group,** Exploratory Research Center on Life and Living Systems (ExCELLS), National Institutes of Natural Sciences, Okazaki, Aichi 444-8585, Japan*

Correspondence to Nobuyasu Koga:*Laboratory for Protein Design, Institute for Protein Research (IPR), Osaka University, Suita, Osaka 565-0871, Japan. *nkoga@protein.osaka-u.ac.jp (N. Koga)*
https://doi.org/10.1016/j.jmb.2024.168791
Edited by Daniel Otzen

Abstract

The vastness of unexplored protein fold universe remains a significant question. Through systematic de novo design of proteins with novel αβ-folds, we demonstrated that nature has only explored a tiny portion of the possible folds. Numerous possible protein folds are still untouched by nature. This review outlines this study and discusses the prospects for design of functional proteins with novel folds.
© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Since the first determination of the three-dimensional structure of myoglobin by Kendrew and colleagues,[1] a vast number of protein structures have been registered in the Protein Data Bank (PDB). These protein structures in the PDB are classified based on sequence and structure, as seen in databases such as CATH,[2] SCOP,[3] and ECOD[4] (the ECOD database focuses more on homology[4]). A principal criterion for the structure-based classification is the fold, also known as topology, which refers to the spatial arrangement of α-helices and β-strands and how they are connected by loops. While many protein structures are experimentally determined each year, the discovery of novel folds has become rare in recent years,[5] although the analyses of the massive structure predictions of naturally occurring proteins by AlphaFold[6,7] have identified additional novel folds.[8–10] Has nature nearly exhausted almost all possible protein folds?

In terms of amino acid sequence space, it is clear that a significant portion of the potential protein universe remains unexplored. Baker illustrates this point with a simple calculation[11]: "For even a relatively small protein of 100 residues, there are $20^{100} = \sim 10^{130}$ possible amino acid sequences (any 1 of the 20 amino acids at each of the 100 positions). For comparison, there are ~10 million species on Earth today, with ~100,000 genes each; $\sim 10^{12}$ proteins in total. The total number of proteins sampled over evolutionary time is likely 10^3–10^5 times larger than this ($\ll 10^{20}$); a tiny fraction of the 10^{130} sequences possible for a 100-residue protein." Considering the vastness of the protein amino acid sequence space, it is plausible that there are more possible folds than those sampled by nature. By exploring beyond what nature has already explored, we may discover a plethora of possible folds.

As with the amino acid sequence space, there are theoretically a large number of folds of αβ-proteins.[12] When considering the αβ-folds by taking into account only the arrangement (i.e., order and orientation) of β-strands within a β-sheet, i.e., β-sheet topology, the total number of theoretical folds can be calculated using the formula $n! \times 2^{n-2}$, where n is the number of β-strands. For β-sheets consisting of three to eight β-strands, there exist a total of 2,754,348 distinct folds. Since this is a theoretical count, it is important to estimate how many of these

0022-2836/© 2024 The Author(s). Published by Elsevier Ltd.This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

folds are actually possible ones, i.e., folds that can actually fold into their stable structures, among the folds that have not yet been sampled by nature. Simply put, how many novel folds exist? We approached this problem by theoretically predicting novel folds and experimentally testing whether it is possible to design proteins that actually fold into those novel folds. While the possibility of the existence of novel folds has been theoretically suggested,[13–15] there have yet been no efforts to explore novel folds through a combination of theory and experiments (i.e., de novo design).

Our approach has been made possible by recent remarkable advancements in de novo protein design, enabling the creation of a diverse array of proteins from scratch.[16] We have developed principles for the design of protein structures stabilized by consistent local and non-local interactions.[17] These principles constitute a set of rules that enabled us to control folds in de novo design of αβ-proteins by selecting the lengths of secondary structures, the lengths or geometries of loops,[18,19] and the registries between β-strands,[20] all optimized for the desired design target fold (Figure 1) (For the design of all-β proteins, rules introduced by Marcos et al. are noted[21]). Utilizing these rules, we sketched out backbone blueprints for de novo designs, built the backbone structures using Rosetta folding simulations[22] based on the blueprints and generated sequences compatible with these backbones using RosettaDesign.[23] For experimental characterization, we selected the designs having funneled energy landscapes, one of the requirements for proteins to be capable of folding,[24] in the Rosetta *ab initio* structure prediction simulations.[25] The experimental results demonstrated that we have successfully created proteins with a variety of naturally observed αβ-folds with atomic accuracy.[18–20] Since our devised rules were also observed in naturally occurring protein structures, it is plausible that natural proteins are stabilized by consistent local and non-local interactions. We also found the proteins designed in this way exhibit high thermal stability, exceeding 100 °C.[26]

Next, to estimate the number of novel folds, we theoretically predicted novel αβ-folds with a three- to eight-stranded β-sheets on a basis of β-sheet topology and systematically carried out computational design and experimental testing on the predicted novel αβ-folds.[12] We attempted to identify potentially foldable αβ-folds by introducing a set of backbone rules for β-sheet topology (Figure 2), which resulted in a total of 12,590 potentially foldable αβ-folds out of the theoretically counted 2,754,348 folds. By selecting the folds that have not been observed in nature, we predicted a total of 12,356 novel αβ-folds with a four- to eight-stranded β-sheet (for four-stranded αβ-folds: 8; five-stranded αβ-folds: 111; six-stranded αβ-folds: 663; seven-stranded αβ-folds: 2,571; eight-stranded αβ-folds: 9,003; for three-stranded αβ-folds, all folds have been observed in nature). To systematically evaluate the feasibility of the predicted novel αβ-folds, we performed de novo design of proteins with all eight predicted novel αβ-folds with a four-stranded β-sheet, including a knot-forming one. Experimental investigation revealed that for all eight predicted novel αβ-folds, the designed proteins folded into the structures as designed (Figure 3). This result demonstrates the ability of the introduced rules to differentiate foldable folds. Given that the designed proteins with all eight predicted novel folds for four-stranded αβ-folds were indeed foldable, it is plausible that proteins with all 12,348 novel αβ-folds with a five- to eight-stranded β-sheet can also fold into their respective fold structures. This number of novel αβ-folds, 12,348, far exceeds the number of αβ-folds observed so far in nature (400 folds). This result suggests that there exist a vast number of

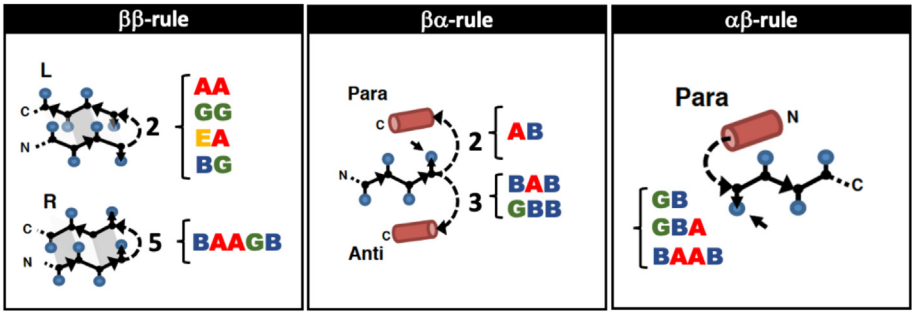

Figure 1. A set of rules for designing protein structures with desired target folds.[18,19] (Left) ββ-rule: The chirality of β-hairpins is controlled by the length or the pattern of backbone geometry of the loop connecting the β-strands. (Middle) βα-rule: the direction from a β-strand to an α-helix, relative to the Cα to Cβ vector at the last strand residue, is controlled by the length or the pattern of backbone geometry of the connecting loop. (Right) αβ-rule: The direction from an α-helix to a β-strand aligns with that of the Cα to Cβ vector at the first strand residue. In all illustrated rules, the preferred lengths and patterns of backbone geometries are detailed. "A" and "B" respectively represent the backbone torsions in the α-helix and β-sheet regions on the Ramachandran plot; G and E correspond to the regions with positive φ values. These rules are used to make backbone blueprints in de novo design of αβ-fold proteins.

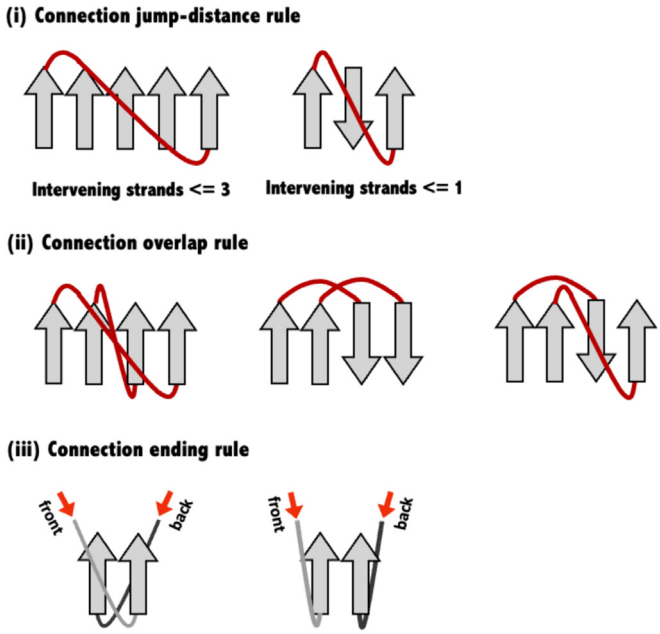

Figure 2. A set of rules for identifying possible folds.[12] (i) Connection jump-distance rule. Left: In *para*-β-*X*-β motifs, where two parallelly aligned β-strands are connected, β-sheets with jump distances (i.e., intervening strands) of three or less are favored. Right: In anti-β-*X*-β motifs, where two anti-parallelly aligned β-strands are connected, β-sheets with jump distances of one or less are favored. *X* represents any backbone conformation. (ii) Connection overlap rule. Geometrical overlap between the connections of two β-*X*-β motifs is unfavorable. (iii) Connection ending rule. When the second strands in two *para*-β-*X*-β motifs are adjacent to each other and aligned parallelly, β-sheets where the two connections end on different sides of the β-sheet are less favored than those where the connections end on the same β-sheet side. The β-sheet arrangements (i.e., folds) satisfying all these three rules are regarded as possible ones.

novel folds that have not yet been explored by nature. A very recent analysis of the structures in AlphaFold Protein Structure Database,[7] which provides over 200 million predicted protein structures, added 96 novel folds into the CATH database[10]: this number is still far smaller than the predicted number of novel folds.

The limitation in the repertories of protein folds found in nature is not due to nature having exhausted all possible protein folds. Rather, the limitation is thought to be a consequence of the evolutionary process. The evolutionary timeline may be insufficient to explore all possible folds, or the limited number of folds at the starting point of life's evolution may have constrained the exploration of novel folds, assuming that the transition from one fold to another novel fold is difficult due to the unlikelihood of intermediates between folds being functional. Alternatively, evolution might have favored adapting and diversifying accidentally discovered folds to specific functions rather than inventing new ones.[27–30] Additionally, the limitation could be due to convergent evolution, that is, particular folds (i.e., superfolds) are favored as a result of convergence to a favorable stereochemical and physical solution.[31] In other words, the size of the amino acid sequence space varies for each fold, folds with a larger sequence space may have appeared more frequently in nature. Furthermore, evolutional pressure to preserve ancestral functional motifs such as the P-loop may have resulted in the limitation of folds.[32,33] In any case, a diverse array of possible novel protein folds are likely to exist. While the study focused on novel αβ-folds, it is probable that many possible folds also exist for both all-α[34] and all-β proteins.

Now is the time to begin inventing functional proteins utilizing novel folds as scaffolds. This effort can provide insight into whether the current protein world we are observing is the only possible solution in terms of structure–function relationship. One of the challenges is whether we can design proteins with novel folds that express functions equivalent to those of naturally occurring proteins. For example, recently, Krishna *et al.*[35] succeeded in designing heme-binding proteins, with α-helical topologies not identified in nature. This achievement raises the question: can we create hemoglobin-like proteins that carry oxygen and possess the allosteric regulation ability, based on novel folds? Such efforts shed light on why the globin-fold was evolutionary chosen to be a fold for generating proteins that transport oxygen with a heme. We

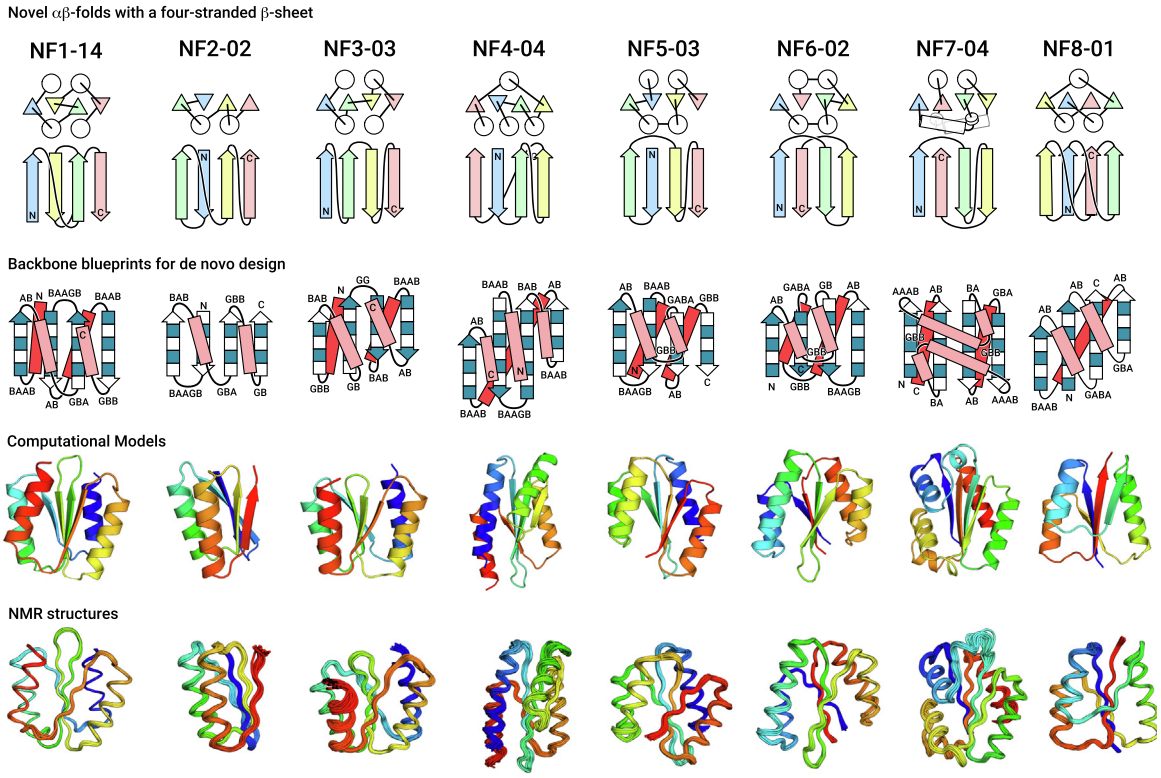

Figure 3. De novo designed proteins with novel αβ-folds.[12] At the top, the first two rows display schematic figures for the predicted eight novel αβ-folds with a four-stranded sheet. Circles and triangles represent α-helix and β-strand, respectively. NF8-01 has a fold forming a knot. In the middle, the backbone blueprints used for the de novo design of the novel αβ-fold structures are shown; these were constructed according to the rules shown in Figure 1. Strand lengths are indicated by filled and open boxes, representing the Cα to Cβ vector coming out of and going into the page, respectively. Letter strings next to the loops indicate their backbone torsion patterns. At the bottom, the two rows present the computational design models and NMR structures.

might be able to create "hemoglobin" in a different and much simpler fold. Another challenge is to design proteins with novel folds that express functions not observed in nature. Naturally occurring protein folds we currently observe may have evolved to customize their structures to perform functions with natural ligands, such as heme, RNA, or DNA: proteins have evolved to suit the needs of life. We humans have created a wide variety of complex artificial compounds that do not exist in nature. For designing proteins that can function with these artificial compounds, novel folds might be suitable.

Attempts at de novo design of proteins began around 1990[36–38] and since then, the technology for protein design has significantly advanced.[39] As a result, it has become possible to design not only protein structures with various shapes but also their functions.[40] Recently developed AI-based methods for de novo design have further accelerated this progress.[35,41–44] In the next decades, by exploring not only the sequence space but also the structural space, beyond what nature has found, various functional proteins will be created.

DATA AVAILABILITY

No data was used for the research described in the article.

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.

Acknowledgements

We thank my colleagues for the collaborations and many discussions on protein design. Especially, we thank D. Baker for the development of de novo design principles. I also thank S. Minami and G. Chikenji for the exploration of novel αβ-folds, and G. Liu, G. T. Montelione, N. Kobayashi and T. Sugiki for NMR structure determination of the designed proteins. We acknowledge support

from the Japan Society for the Promotion of Science (JSPS) KAKENHI Grants-in-Aid for Scientific Research 24H02261.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author used ChatGPT in order to improve readability. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Received 2 April 2024;
Accepted 5 September 2024;
Available online 12 September 2024

Keywords:
protein design;
protein fold;
novel fold;
protein sequence space;
protein topology

References

1. Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., Phillips, D.C., (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662–666.
2. Orengo, C.A., Jones, D.T., Thornton, J.M., (1994). Protein superfamilies and domain superfolds. *Nature* **372**, 631–634.
3. Murzin, A.G., Brenner, S.E., Hubbard, T., Chothia, C., (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540.
4. Cheng, H., Schaeffer, R.D., Liao, Y., Kinch, L.N., Pei, J., Shi, S., Kim, B.-H., Grishin, N.V., (2014). ECOD: an evolutionary classification of protein domains. *PLoS Comput. Biol.* **10**, e1003926.
5. Cuff, A.L., Sillitoe, I., Lewis, T., Redfern, O.C., Garratt, R., Thornton, J., Orengo, C.A., (2009). The CATH classification revisited–architectures reviewed and new ways to characterize structural divergence in superfamilies. *Nucleic Acids Res.* **37**, D310–D314.
6. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589.
7. Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., Tsenkov, M., Nair, S., Mirdita, M., Yeo, J., Kovalevskiy, O., Tunyasuvunakool, K., Laydon, A., Žídek, A., Tomlinson, H., Hariharan, D., Abrahamson, J., Green, T., Jumper, J., Birney, E., Steinegger, M., Hassabis, D., Velankar, S., (2024). AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* **52**, D368–D375.
8. Bordin, N., Sillitoe, I., Nallapareddy, V., Rauer, C., Lam, S. D., Waman, V.P., Sen, N., Heinzinger, M., Littmann, M., Kim, S., Velankar, S., Steinegger, M., Rost, B., Orengo, C., (2023). AlphaFold2 reveals commonalities and novelties in protein structure space for 21 model organisms. *Commun. Biol.* **6**, 160.
9. Durairaj, J., Waterhouse, A.M., Mets, T., Brodiazhenko, T., Abdullah, M., Studer, G., Tauriello, G., Akdel, M., Andreeva, A., Bateman, A., Tenson, T., Hauryliuk, V., Schwede, T., Pereira, J., (2023). Uncovering new families and folds in the natural protein universe. *Nature* **622**, 646–653.
10. Waman, V.P., Bordin, N., Alcraft, R., Vickerstaff, R., Rauer, C., Chan, Q., Sillitoe, I., Yamamori, H., Orengo, C., (2024). CATH 2024: CATH-AlphaFlow doubles the number of structures in CATH and reveals nearly 200 new folds. *J. Mol. Biol.*, 168551.
11. Baker, D., (2019). What has de novo protein design taught us about protein folding and biophysics? *Protein Sci.* **28**, 678–683.
12. Minami, S., Kobayashi, N., Sugiki, T., Nagashima, T., Fujiwara, T., Tatsumi-Koga, R., Chikenji, G., Koga, N., (2023). Exploration of novel αβ-protein folds through de novo design. *Nature Struct. Mol. Biol.* **30**, 1132–1140.
13. 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.
14. Cossio, P., Trovato, A., Pietrucci, F., Seno, F., Maritan, A., Laio, A., (2010). Exploring the universe of protein structures beyond the Protein Data Bank. *PLoS Comput. Biol.* **6**, e1000957.
15. Chitturi, B., Shi, S., Kinch, L.N., Grishin, N.V., (2016). Compact structure patterns in proteins. *J. Mol. Biol.* **428**, 4392–4412.
16. Kortemme, T., (2024). De novo protein design—from new structures to programmable functions. *Cell* **187**, 526–544.
17. Koga, R., Koga, N., (2019). Consistency principle for protein design. *Biophys. Physicobiol.* **16**, 304–309.
18. Koga, N., Tatsumi-Koga, R., Liu, G., Xiao, R., Acton, T.B., Montelione, G.T., Baker, D., (2012). Principles for designing ideal protein structures. *Nature* **491**, 222–227.
19. Lin, Y.-R., Koga, N., Tatsumi-Koga, R., Liu, G., Clouser, A. F., Montelione, G.T., Baker, D., (2015). Control over overall shape and size in de novo designed proteins. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E5478–E5485.
20. Koga, N., Koga, R., Liu, G., Castellanos, J., Montelione, G. T., Baker, D., (2021). Role of backbone strain in de novo design of complex α/β protein structures. *Nature Commun.* **12**, 3921.
21. Marcos, E., Chidyausiku, T.M., McShan, A.C., Evangelidis, T., Nerli, S., Carter, L., Nivón, L.G., Davis, A., Oberdorfer, G., Tripsianes, K., Sgourakis, N.G., Baker, D., (2018). De novo design of a non-local β-sheet protein with high stability and accuracy. *Nature Struct. Mol. Biol.* **25**, 1028–1034.

22. Simons, K.T., Kooperberg, C., Huang, E., Baker, D., (1997). Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* **268**, 209–225.
23. 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.
24. Leopold, P., Montal, M., Onuchic, J., (1992). Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 8721–8725.
25. Rohl, C.A., Strauss, C.E.M., Misura, K.M.S., Baker, D., (2004). Protein structure prediction using Rosetta. *Methods Enzymol.* **383**, 66–93.
26. Koga, R., Yamamoto, M., Kosugi, T., Kobayashi, N., Sugiki, T., Fujiwara, T., Koga, N., (2020). Robust folding of a de novo designed ideal protein even with most of the core mutated to valine. *Proc. Natl. Acad. Sci.* **117**, 31149–31156.
27. Levitt, M., (2009). Nature of the protein universe. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 11079–11084.
28. Bornberg-Bauer, E., Albà, M.M., (2013). Dynamics and adaptive benefits of modular protein evolution. *Curr. Opin. Struct. Biol.* **23**, 459–466.
29. Rost, B., (2002). Did evolution leap to create the protein universe? *Curr. Opin. Struct. Biol.* **12**, 409–416.
30. Todd, A.E., Orengo, C.A., Thornton, J.M., (2001). Evolution of function in protein superfamilies, from a structural perspective. *J. Mol. Biol.* **307**, 1113–1143.
31. Orengo, C.A., Flores, T.P., Jones, D.T., Taylor, W.R., Thornton, J.M., (1993). Recurring structural motifs in proteins with different functions. *Curr. Biol.* **3**, 131–139.
32. Alva, V., Söding, J., Lupas, A.N., (2015). A vocabulary of ancient peptides at the origin of folded proteins. *Elife* **4**, e09410.
33. Longo, L.M., Jabłońska, J., Vyas, P., Kanade, M., Kolodny, R., Ben-Tal, N., Tawfik, D.S., (2020). On the emergence of P-Loop NTPase and Rossmann enzymes from a Beta-Alpha-Beta ancestral fragment. *Elife* **9**
34. Sakuma, K., Kobayashi, N., Sugiki, T., Nagashima, T., Fujiwara, T., Suzuki, K., Kobayashi, N., Murata, T., Kosugi, T., Tatsumi-Koga, R., Koga, N., (2024). Design of complicated all-α protein structures. *Nature Struct. Mol. Biol.* **31**, 275–282.
35. Krishna, R., Wang, J., Ahern, W., Sturmfels, P., Venkatesh, P., Kalvet, I., Lee, G.R., Morey-Burrows, F.S., Anishchenko, I., Humphreys, I.R., McHugh, R., Vafeados, D., Li, X., Sutherland, G.A., Hitchcock, A., Hunter, C.N., Kang, A., Brackenbrough, E., Bera, A.K., Baek, M., DiMaio, F., Baker, D., (2024). Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* **384**, eadl2528.
36. Ho, S.P., DeGrado, W.F., (1987). Design of a 4-helix bundle protein: synthesis of peptides which self-associate into a helical protein. *J. Am. Chem. Soc.* **109**, 6751–6758.
37. Hecht, M.H., Richardson, J.S., Richardson, D.C., Ogden, R.C., (1990). De novo design, expression, and characterization of Felix: a four-helix bundle protein of native-like sequence. *Science* **249**, 884–891.
38. Woolfson, D.N., (2021). A brief history of de Novo protein design: minimal, rational, and computational. *J. Mol. Biol.* **433**, 167160.
39. Pan, X., Kortemme, T., (2021). Recent advances in de Novo protein design: principles, methods, and applications. *J. Biol. Chem.* **296**, 100558.
40. Chu, A.E., Lu, T., Huang, P.-S., (2024). Sparks of function by de novo protein design. *Nature Biotechnol.* **42**, 203–215.
41. Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R.J., Milles, L.F., Wicky, B.I.M., Courbet, A., de Haas, R.J., Bethel, N., Leung, P.J.Y., Huddy, T.F., Pellock, S., Tischer, D., Chan, F., Koepnick, B., Nguyen, H., Kang, A., Sankaran, B., Bera, A.K., King, N.P., Baker, D., (2022). Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **378**, 49–56.
42. Watson, J.L., Juergens, D., Bennett, N.R., Trippe, B.L., Yim, J., Eisenach, H.E., Ahern, W., Borst, A.J., Ragotte, R. J., Milles, L.F., Wicky, B.I.M., Hanikel, N., Pellock, S.J., Courbet, A., Sheffler, W., Wang, J., Venkatesh, P., Sappington, I., Torres, S.V., Lauko, A., De Bortoli, V., Mathieu, E., Ovchinnikov, S., Barzilay, R., Jaakkola, T.S., DiMaio, F., Baek, M., Baker, D., (2023). De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100.
43. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., Bodenstein, S.W., Evans, D.A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A.I., Cowie, A., Figurnov, M., Fuchs, F.B., Gladman, H., Jain, R., Khan, Y.A., Low, C.M.R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E.D., Zielinski, M., Žídek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., Jumper, J.M., (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500.
44. Ingraham, J.B., Baranov, M., Costello, Z., Barber, K.W., Wang, W., Ismail, A., Frappier, V., Lord, D.M., Ng-Thow-Hing, C., Van Vlack, E.R., Tie, S., Xue, V., Cowles, S.C., Leung, A., Rodrigues, J.V., Morales-Perez, C.L., Ayoub, A. M., Green, R., Puentes, K., Oplinger, F., Panwar, N.V., Obermeyer, F., Root, A.R., Beam, A.L., Poelwijk, F.J., Grigoryan, G., (2023). Illuminating protein space with a programmable generative model. *Nature* **623**, 1070–1078.