Special Issue
"Progress of Theoretical and Computational Biophysics"

Vol. 16, pp. 304–309 (2019)
doi: 10.2142/biophysico.16.0_304

Biophysics and Physicobiology

https://www.jstage.jst.go.jp/browse/biophysico/

Review Article

Consistency principle for protein design

Rie Koga[1] and Nobuyasu Koga[1,2,3]

[1]*Protein Design Group, Exploratory Research Center on Life and Living Systems (ExCELLS), National Institutes of Natural Sciences, Okazaki, Aichi 444-8585, Japan*
[2]*Research Center of Integrative Molecular Systems, Institute for Molecular Science, National Institutes of Natural Sciences, Okazaki, Aichi 444-8585, Japan*
[3]*SOKENDAI, The Graduate University for Advanced Studies, Shonan Village, Hayama, Kanagawa 240-0193, Japan*

Received June 28, 2019; accepted August 1, 2019

Protein design holds promise for applications such as the control of cells, therapeutics, new enzymes and protein-based materials. Recently, there has been progress in rational design of protein molecules, and a lot of attempts have been made to create proteins with functions of our interests. The key to the progress is the development of methods for controlling desired protein tertiary structures with atomic-level accuracy. A theory for protein folding, the consistency principle, proposed by Nobuhiro Go in 1983, was a compass for the development. Anfinsen hypothesized that proteins fold into the free energy minimum structures, but Go further considered that local and non-local interactions in the free energy minimum structures are consistent with each other. Guided by the principle, we proposed a set of rules for designing ideal protein structures stabilized by consistent local and non-local interactions. The rules made possible designs of amino acid sequences with funnel-shaped energy landscapes toward our desired target structures. So far, various protein structures have been created using the rules, which demonstrates significance of our rules as intended. In this review, we briefly describe how the consistency principle impacts on our efforts for developing the design technology.

Key words: computational protein design, ideal proteins, funnel-shaped energy landscapes, consistent local and non-local interactions, Go model

Corresponding authors: Rie Koga, Protein Design Group, Exploratory Research Center on Life and Living Systems, National Institutes of Natural Sciences, 38 Nishigo-Naka, Myodaiji, Okazaki, Aichi 444-8585, Japan. e-mail: riekoga@ims.ac.jp; Nobuyasu Koga, Protein Design Group, Exploratory Research Center on Life and Living Systems, National Institutes of Natural Sciences, 38 Nishigo-Naka, Myodaiji, Okazaki, Aichi 444-8585, Japan. e-mail: nkoga@ims.ac.jp

Understanding of protein folding is important to develop the methodology for creating our desired proteins. Anfinsen hypothesized that proteins fold into the free energy minimum structures [1]. However, the folding problem—How do amino acid sequences determine the folded structures?—has been a long-standing problem for more than a half century. Researches for protein folding or structure prediction from amino acid sequences have attempted to address the problem by studying complicated proteins created by nature spending billions of years, which have energetically unfavorable non-ideal features such as kinked α-helices, bulged β-strands and buried polar residues. Protein design studies provide an alternative approach to tackle the problem by creating simple protein structures not having such unfavorable features from scratch with hypotheses about protein folding and experimentally testing how the designs fold.

◀ ***Significance*** ▶

Protein design expands the possibility of developments for therapeutics, biosensors, materials, etc. Recently there has been great progress in computational design of protein structures. The basic idea that underlies the progress is the rules we discovered relating secondary structure patterns to tertiary motifs, which make it possible to design Go's proposed ideal protein structures. In this review, we describe how our rules were discovered in the history of protein design and folding studies.

©2019 THE BIOPHYSICAL SOCIETY OF JAPAN

Protein design in early days

Protein design work was started in the late 1980s by the design of helical bundle structures. DeGrado, W. H., *et al.* [2] attempted to design dimeric helical bundle structures based on a hydrophobic (H) and hydrophilic (P) amino acid sequence pattern using the helical wheel model. Hecht, M. H., *et al.* [3] tried to design a four helix bundle, taking into account various structural features for helical proteins, derived from statistics of known protein structures, such as favorable amino acid types in α-helix or at the N- and C-terminal α-helix capping positions. However, these designed proteins were experimentally found to be in a molten globule state, in which proteins are compact with native-like secondary structure contents but without tight core packing [4]. The designs of TIM-barrel fold were also challenged by considering sequence preference for each residue position based on natural TIM-barrel proteins, but also found to be in a molten globule state [5]. All these efforts in early days tell us that protein folding is not determined just by simple amino acid patterning such as HP pattern: detailed atomistic modeling of tertiary structures is essential for designing amino acid sequences that have folding ability toward a unique tertiary structure. For example, the core in natural protein structures is apparently made of densely packed hydrophobic atoms. Actually, Hecht, M. H., *et al.* [3] tried to achieve the packed hydrophobic core in the design using a physical model, but it would have been difficult to capture the atomistic detail by hand.

Computational protein design from sidechain-redesign to full-scratch

The pioneer work of computational protein design with atomic resolution modeling was done by Dahiyat, B. I., *et al.* in the late 1990s, focusing on the redesign of sidechains of a naturally occurring protein structure using the backbone as a scaffold [6]. The group redesigned sidechains of zinc finger domain by stripping off the native sidechains of the protein and rebuilding new sidechains (amino acids) with a set of discretely represented sidechain conformations (rotamer library): new sidechain conformations that have the lowest energy for the zinc finger backbone were explored with the rotamer library. The design was found to form a compact well-ordered structure of zinc finger domain in the solution NMR structures, in which the packing of the hydrophobic core was similar to the design model. Since then, successful sidechain-redesigns of natural proteins have been reported for lambda-Cro [7], tenascin [8], homeodomain [9], etc. In these days, various designs such as novel enzymes [10], an influenza binder [11], and cage-like symmetric oligomers [12–15] were created using naturally occurring proteins as scaffolds, which can be considered as applications of the rotamer-based sidechain designs.

The first de novo design of a globular protein structure was achieved by Kuhlman, B., *et al.* in 2003 [16]. The authors created a novel protein fold, Top7, from scratch (Fig. 4). In this study, the authors developed a protocol in the software, Rosetta, to design protein structures from the backbone, in which the backbone structure of Top7 was built by assembling short fragments of known protein structures [17] and then sidechains that stabilize the built backbone were explored by an iterative approach between the rotamer-based sidechain design for a fixed backbone and gradient-based optimization of the entire structure for a fixed sequence [16]. The developed protocol enabled to identify sidechain-backbone pairs that have very low energies in computation, and one of the sequences of the pairs was found to have folding ability to the designed Top7 structure with atomic-level accuracy. Since then, however, no one made a success of the de novo design of protein structures until our work, indicating that the protocol was not enough to design protein structures. For designing proteins folded into the desired structures, there should be other essential factors than exploring low energy structures with tight hydrophobic core packing. Indeed, the paper did not describe how the lengths of the secondary structures and loops in the Top7 structure were determined. If we can design proteins with well-packed low energy structures whatever lengths of the secondary structures and loops are used, are the designs foldable? We started our work [18] in Baker group by investigating the folding abilities of proteins depending on the lengths of secondary structures and loops with folding simulations and statistical analysis of naturally occurring protein structures. Before describing the work, some hypotheses suggested by protein folding studies need to be introduced, which were the compass for developing our design methods.

Funnel-shaped energy landscape

The theoretical studies for protein folding from the late 1980s to 1990s suggested a hypothesis that natural proteins have evolved to have funnel-shaped energy landscapes toward the native state from the denatured state, in which proteins decrease their energies along with the formation of the folded structure [19–22] (Fig. 1a). On the other hand, polypeptides with random amino acid sequences have various low energy structures, resulting in rugged and non-funneled energy landscapes (Fig. 1b). Such polypeptides are trapped into various non-native states, not showing foldability toward the native state. This funnel hypothesis is supported by the fact that protein folding studies using Go-like models can explain the folding mechanisms (cooperative folding-unfolding transition, pathways, rates, etc.) found by experiments for small proteins [23–28]. The original Go model (not Go-like), a lattice model, proposed by Go [29] to embody the consistency principle [30] was applied to study the cooperative protein folding-unfolding transition, not in the context of the above energy landscape discussion. In either case of the Go or Go-like models, the essence of the

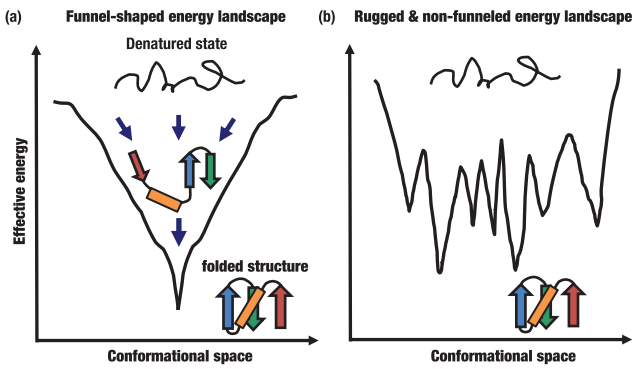

Figure 1 Schematic of (a) a funnel-shaped energy landscape of an amino acid sequence that has folding ability toward a folded structure and (b) a rugged and non-funneled energy landscape of a random sequence.

models is an assumption to consider only native interactions formed in the native conformation as the energy gain ignoring non-native interactions, which makes the energy landscape smoothly funneled.

The funnel theory directly leads to the idea that for designing foldable proteins, it is essential to obtain amino acid sequences with funnel-shaped energy landscapes toward target structures. Such sequences could be acquired by exploring sequences that simultaneously stabilize the target structure and destabilize all of non-native structures. However, it is practically impossible to identify such sequences considering myriad non-native structures. How do we then find out such sequences? Clues for solving the question were the consistency principle proposed by Nobuhiro Go in 1983 [30] and the discussion by Chikenji, G., *et al.* on how funnel-shaped energy landscapes arise [31].

Go's consistency principle and local backbone preference for shaping funnel

Go proposed a hypothesis for protein folding, the consistency principle [30]. He considered that the local and non-local interactions are consistent with each other, in which local (non-local) interactions are those between amino acids close (distant) along a sequence. For example, if a tertiary structure is stabilized by non-local interactions such as hydrophobic and vdW interactions, but has local steric clashes or amino acids with low propensity for their secondary structures, the interactions of the tertiary structure are regarded as inconsistent. For the consistent case, all the local and non-local interactions consistently stabilize the tertiary structure with each other. The Go's consistency principle is a paradigm shift for protein folding next to the Anfinsen's thermodynamic principle: a folded structure has consistent interactions as well as corresponding to a free energy minimum. Indeed, as of 2019, more than 150,000 structures are deposited in the protein structure database (PDB), and when observing the structures, various interactions are surprisingly consistent: for example, buried polar groups without making hydrogen bonds are very rare [32]. About 20 years after the Go's consistency principle, Chikenji, G., *et al.* discussed how funnel-shaped energy landscapes arise, using exact enumeration with a HP lattice model [31]. As described above, the Go model considers only the specific interactions formed in the native structure as an ideal limit to satisfy the consistency principle and has smooth-funneled energy landscape. On the other hand, the HP model used in their study has nonspecific nonlocal hydrophobic interactions and has a rugged energy landscape with multiple stable non-native structures. The authors demonstrated that by introducing local interactions into the HP model through a prohibition of one conformation for each local sequence (Fig. 2b), the rugged energy landscape get sculpted to be funneled into the native structure, in which any disallowed local conforma-

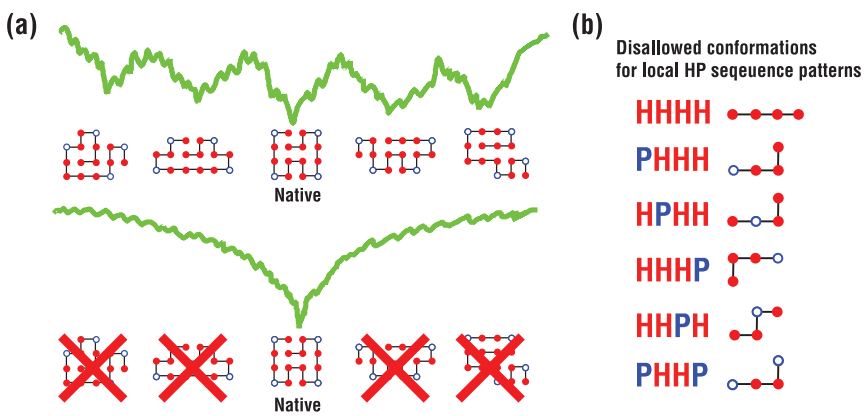

Figure 2 Illustration for how funnel-shaped energy landscapes arise using lattice HP model by Chikenji, G., *et al.* [31]. (a) The HP sequence, HHHPHHPHHHHPHHPH, which has initially a rugged energy landscape with multiple low-energy conformations, becomes to have a funnel-shaped energy landscape toward a single conformation (Native) by assuming (b) just single disallowed conformation for each local sequence pattern.

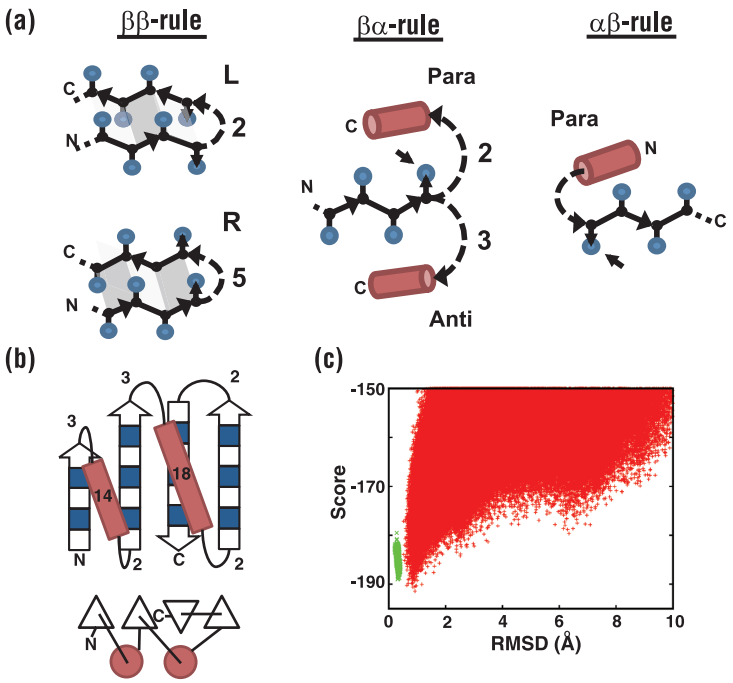

Figure 3 (a) Discovered rules for designing ideal protein structures stabilized consistent local and non-local interactions. (b) A blueprint, drawn by the rules, for building backbone structures for an ideal structure of IF3-like fold shown in Figure 4. The numbers represent the secondary structure and loop lengths. Strand lengths are shown by filled and open boxes. The filled boxes represent pleats (Cα-Cβ vectors) coming out of the page, and the open boxes represent pleats going into the page. (c) The energy landscape for the designed sequence of IF3-like fold shown in Figure 4. The energy landscape was obtained from Rosetta ab initio structure prediction simulations [33]. Each red point represents the lowest energy structure obtained in independent simulation starting from an extended chain; the y-axis shows the Rosetta all atom energy, and the x-axis, the Cα root mean square deviation (RMSD) to the design model. Each green point represents the lowest energy structure obtained in independent simulation starting from the design model.

tions are not included and maximum number of hydrophobic interactions is satisfied, i.e., local and nonlocal interactions are consistent (Fig. 2a). This result suggested that the conformational biases by local interactions can shape funnel-shaped energy landscapes.

Based on the described studies on protein folding, we have sought to develop methods for designing protein structures from scratch. Go indicated that the interactions in naturally occurring proteins (real proteins) cannot be perfectly consistent because interactions relating to stabilities and functions may not be consistent. Moreover, he suggested a concept of ideal proteins, in which various interactions are perfectly consistent. We hypothesized that proteins with funnel-shaped energy landscapes can be readily generated by designing such ideal proteins, and set out to seek design methods to create them [18].

The rules for designing ideal proteins

We investigated the relationships between local interactions favoring secondary structure patterns and non-local interactions favoring tertiary structure motifs [18] using Rosetta folding simulations [33] and statistical analysis of naturally occurring protein structures. As the results, we found that folding ability to a particular tertiary motif is strongly dependent on the lengths of the secondary structures and the connecting loop, not the detail of amino acid sequences, and that these dependencies are described in the simple rules (Fig. 3a). Our succeeding paper [34] further identified that the rules can be extended with the discretized backbone torsion bins, ABEGO, for the loops (torsion bins A and B are the α-helix and β-sheet regions; G and E are the positive phi regions equivalent to A and B; and O is the cis peptide bond). The major origin of the backbone structure preferences found in the rules is backbone strain arising from the polypeptide's molecular geometry and the local steric hindrance in phi-psi angles of each residue. The discovered rules allow us to control protein topologies: selecting lengths or ABEGO patterns of the secondary structures and loops that favor the tertiary motifs present in the desired topology, many of the non-native topologies are disfavored by local backbone strain, resulting in funnel-shaped energy landscape.

Design of various ideal protein structures based on the rules

We have finally reached a general method to design amino acid sequences with funnel-shaped energy landscapes toward a unique structure [18]. De novo protein designs proceed in two steps: the backbone building [17] and the side-

chain building (sequence-design) that stabilizes the backbone [16]. Amino acid sequences with funnel-shaped energy landscapes can be readily designed by building the backbone structures with a blueprint (Fig. 3b), in which the lengths or ABEGO patterns of the secondary structures and loops are determined so that the tertiary motifs present in the target topology are favored using the discovered rules, and then by designing sidechains that simultaneously stabilize the local secondary structures and the non-local tertiary structures. Our developed design principles made possible the designs of various ideal protein structures stabilized by consistent local and non-local interactions with strongly funneled energy landscapes (Fig. 3c). We succeeded in designing αβ-protein structures with five different topologies *de novo* (Fig. 4, the top and middle rows except Top7) [18] as well as those with the same topologies but different size and shape [34]. The design of TIM-barrel fold was also finally achieved by Huang, P. S., *et al.* using our design principles (Fig. 4) [35]. Notably, Top7 can be considered as one of the ideal proteins: the lengths of secondary structures and loops of Top7 completely agree with our rules [16]. It is quite surprising that Kuhlman, B., *et al.* had selected the appropriate lengths without the rules. The concept for designing protein structures by using the relationships between secondary structure patterns and tertiary structures was further applied to the design of all-α or all-β proteins. Marcos, E., *et al.* identified the rules for β-arches, in which loops connect two β-strands belonging to different β-sheets, and succeeded in designing a non-local β-sheet protein, jellyroll structure, *de novo* (Fig. 4) [36]. Likewise, Doyle, L., *et al.* and Brunette, T. J., *et al.* designed α-helical tandem repeat proteins successfully *de novo* (Fig. 4) [37,38].

Conclusion

How do amino acid sequences determine the folded structures? Complicated tertiary structures of naturally occurring proteins have obscured the principles for protein folding, but now one end of those was revealed through our design studies seeking Go's ideal proteins: the funnel-shaped energy landscapes of naturally occurring proteins probably emerged as the results of destabilization of the myriad non-native structures through the stabilization of local structures that disfavor non-native topologies. Protein is sometimes likened to a string and loops are considered very flexible, but this is not a correct understanding. One of the appropriate examples for proteins would be snake cube model, in which the tertiary shapes are limited by local restraint (See Go's article in this volume). Now, we are ready to explore enormously large space of protein structure universe with the rules for creating novel proteins of our interests. The challenges have already started.

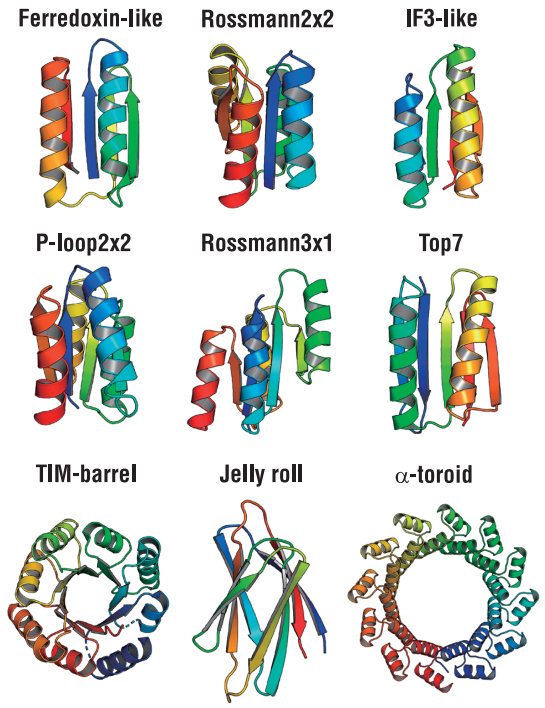

Figure 4 Examples of designed ideal protein structures based on the rules. Ferredoxin-like, Rossmann2x2, IF3-like, P-loop2x2, Rossmann3x1 were designed by Koga, N., *et al.* in 2012 [18]: Top7, by Kuhlman, B., *et al.* in 2003 [16]: TIM-barrel, by Huang, P. S., *et al.* in 2016 [35]: Jelly roll, by Marcos, E., *et al.* in 2018 [36]: α-toroid, by Doyle, L., *et al.* in 2015 [37]. Experimentally determined structures by NMR or X-ray were used for drawing.

Acknowledgements

We gratefully thank Nobuhiro Go, a pioneer for protein folding studies. His proposed concept, the consistency principle, has been center of our research. Our protein design work was done in Baker Group. We thank David Baker and all members in his group. We also thank all researchers who had been and have been working for protein folding and design.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

R. K and N. K reviewed the protein design and folding studies and wrote the manuscript.

References

[1] Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).
[2] DeGrado, W. F., Regan, L. & Ho, S. P. The design of a four-helix bundle protein. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 521–526 (1987).

[3] Hecht, M. H., Richardson, J. S., Richardson, D. C. & Ogden, R. C. De novo design, expression, and characterization of Felix: a four-helix bundle protein of native-like sequence. *Science* **249**, 884–891 (1990).
[4] Ohgushi, M. & Wada, A. 'Molten-globule state': a compact form of globular proteins with mobile side-chains. *FEBS Lett.* **164**, 21–24 (1983).
[5] Tanaka, T., Hayashi, M., Kimura, H., Oobatake, M. & Nakamura, H. De novo design and creation of a stable artificial protein. *Biophys. Chem.* **50**, 47–61 (1994).
[6] Dahiyat, B. I. & Mayo, S. L. De novo protein design: fully automated sequence selection. *Science* **278**, 82–87 (1997).
[7] Isogai, Y., Ito, Y., Ikeya, T., Shiro, Y. & Ota, M. Design of lambda Cro fold: solution structure of a monomeric variant of the de novo protein. *J. Mol. Biol.* **354**, 801–814 (2005).
[8] Hu, X., Wang, H., Ke, H. & Kuhlman, B. Computer-based redesign of a beta sandwich protein suggests that extensive negative design is not required for de novo beta sheet design. *Structure* **16**, 1799–1805 (2008).
[9] Shah, P. S., Hom, G. K., Ross, S. A., Lassila, J. K., Crowhurst, K. A. & Mayo, S. L. Full-sequence computational design and solution structure of a thermostable protein variant. *J. Mol. Biol.* **372**, 1–6 (2007).
[10] Mak, W. S. & Siegel, J. B. Computational enzyme design: transitioning from catalytic proteins to enzymes. *Curr. Opin. Struct. Biol.* **27**, 87–94 (2014).
[11] Fleishman, S. J., Whitehead, T. A., Ekiert, D. C., Dreyfus, C., Corn, J. E., Strauch, E. M., *et al.* Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **332**, 816–821 (2011).
[12] King, N. P., Sheffler, W., Sawaya, M. R., Vollmar, B. S., Sumida, J. P., Andre, I., *et al.* Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336**, 1171–1174 (2012).
[13] King, N. P., Bale, J. B., Sheffler, W., McNamara, D. E., Gonen, S., Gonen, T., *et al.* Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103–108 (2014).
[14] Bale, J. B., Gonen, S., Liu, Y., Sheffler, W., Ellis, D., Thomas, C., *et al.* Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389–394 (2016).
[15] Butterfield, G. L., Lajoie, M. J., Gustafson, H. H., Sellers, D. L., Nattermann, U., Ellis, D., *et al.* Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature* **552**, 415–420 (2017).
[16] Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L. & Baker, D. Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).
[17] Simons, K. T., Kooperberg, C., Huang, E. & Baker, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* **268**, 209–225 (1997).
[18] Koga, N., Tatsumi-Koga, R., Liu, G., Xiao, R., Acton, T. B., Montelione, G. T., *et al.* Principles for designing ideal protein structures. *Nature* **491**, 222–227 (2012).
[19] Bryngelson, J. D. & Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA* **84**, 7524–7528 (1987).
[20] Leopold, P. E., Montal, M. & Onuchic, J. N. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA* **89**, 8721–8725 (1992).
[21] Shakhnovich, E. I. & Gutin, A. M. Engineering of stable and fast-folding sequences of model proteins. *Proc. Natl. Acad. Sci. USA* **90**, 7195–7199 (1993).
[22] Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* **48**, 545–600 (1997).
[23] Galzitskaya, O. V. & Finkelstein, A. V. A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proc. Natl. Acad. Sci. USA* **96**, 11299–11304 (1999).
[24] Alm, E. & Baker, D. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc. Natl. Acad. Sci. USA* **96**, 11305–11310 (1999).
[25] Munoz, V. & Eaton, W. A. A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc. Natl. Acad. Sci. USA* **96**, 11311–11316 (1999).
[26] Clementi, C., Jennings, P. A. & Onuchic, J. N. How native-state topology affects the folding of dihydrofolate reductase and interleukin-1beta. *Proc. Natl. Acad. Sci. USA* **97**, 5871–5876 (2000).
[27] Clementi, C., Nymeyer, H. & Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **298**, 937–953 (2000).
[28] Koga, N. & Takada, S. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. *J. Mol. Biol.* **313**, 171–180 (2001).
[29] Taketomi, H., Ueda, Y. & Go, N. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* **7**, 445–459 (1975).
[30] Go, N. Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.* **12**, 183–210 (1983).
[31] Chikenji, G., Fujitsuka, Y. & Takada, S. Shaping up the protein folding funnel by local interaction: lesson from a structure prediction study. *Proc. Natl. Acad. Sci. USA* **103**, 3141–3146 (2006).
[32] Fleming, P. J. & Rose, G. D. Do all backbone polar groups in proteins form hydrogen bonds? *Protein Sci.* **14**, 1911–1917 (2005).
[33] Rohl, C. A., Strauss, C. E., Misura, K. M. & Baker, D. Protein structure prediction using Rosetta. *Methods Enzymol.* **383**, 66–93 (2004).
[34] Lin, Y. R., Koga, N., Tatsumi-Koga, R., Liu, G., Clouser, A. F., Montelione, G. T., *et al.* Control over overall shape and size in de novo designed proteins. *Proc. Natl. Acad. Sci. USA* **112**, E5478–5485 (2015).
[35] Huang, P. S., Feldmeier, K., Parmeggiani, F., Velasco, D. A. F., Hocker, B. & Baker, D. De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat. Chem. Biol.* **12**, 29–34 (2016).
[36] Marcos, E., Chidyausiku, T. M., McShan, A. C., Evangelidis, T., Nerli, S., Carter, L., *et al.* De novo design of a non-local β-sheet protein with high stability and accuracy. *Nat. Struct. Mol. Biol.* **25**, 1028–1034 (2018).
[37] Doyle, L., Hallinan, J., Bolduc, J., Parmeggiani, F., Baker, D., Stoddard, B. L., *et al.* Rational design of α-helical tandem repeat proteins with closed architectures. *Nature* **528**, 585–588 (2015).
[38] Brunette, T. J., Parmeggiani, F., Huang, P. S., Bhabha, G., Ekiert, D. C., Tsutakawa, S. E., *et al.* Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).

This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. To view a copy of this license, visit https://creativecommons.org/licenses/by-nc-sa/4.0/.