Robust folding of a de novo designed ideal protein even with most of the core mutated to valine

Rie Koga[a,1], Mami Yamamoto[b,2], Takahiro Kosugi[a,b,c,2], Naohiro Kobayashi[d,e,2], Toshihiko Sugiki[d], Toshimichi Fujiwara[d], and Nobuyasu Koga[a,b,c,1]

[a]Protein Design Group, Exploratory Research Center on Life and Living Systems, National Institutes of Natural Sciences, Okazaki, Aichi 444-8585, Japan; [b]Department of Structural Molecular Science, School of Physical Sciences, The Graduate University for Advanced Studies, SOKENDAI, Hayama, Kanagawa 240-0193, Japan; [c]Research Center of Integrative Molecular Systems, Institute for Molecular Science, National Institutes of Natural Sciences, Okazaki, Aichi 444-8585, Japan; [d]Laboratory of Molecular Biophysics, Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan; and [e]RIKEN Spring-8 Center, RIKEN, Yokohama, Kanagawa 230-0045, Japan

Edited by William F. DeGrado, University of California, San Francisco, CA, and approved October 21, 2020 (received for review February 6, 2020)

Protein design provides a stringent test for our understanding of protein folding. We previously described principles for designing ideal protein structures stabilized by consistent local and nonlocal interactions, based on a set of rules relating local backbone structures to tertiary packing motifs. The principles have made possible the design of protein structures having various topologies with high thermal stability. Whereas nonlocal interactions such as tight hydrophobic core packing have traditionally been considered to be crucial for protein folding and stability, the rules proposed by our previous studies suggest the importance of local backbone structures to protein folding. In this study, we investigated the robustness of folding of de novo designed proteins to the reduction of the hydrophobic core, by extensive mutation of large hydrophobic residues (Leu, Ile) to smaller ones (Val) for one of the designs. Surprisingly, even after 10 Leu and Ile residues were mutated to Val, this mutant with the core mostly filled with Val was found to not be in a molten globule state and fold into the same backbone structure as the original design, with high stability. These results indicate the importance of local backbone structures to the folding ability and high thermal stability of designed proteins and suggest a method for engineering thermally stabilized natural proteins.

de novo protein design | protein folding | thermal stability | local backbone structures

The de novo design of protein structures, starting from pioneering work (1, 2), has been achieved in tandem with our understanding of how amino acid sequences determine folded structures (3–16). A breakthrough in protein design methodology was a finding of principles for encoding funnel-shaped energy landscapes into amino acid sequences (7, 10, 17, 18). Based on studies of protein folding, it had been suggested that naturally occurring proteins have evolved to have funnel-shaped energy landscapes toward their folded structures (19–23). However, complicated structures of naturally occurring proteins with nonideal features for folding—for example, kinked α-helices, bulged β-strands, long or strained loops, and buried polar groups—make it difficult to understand how the funnels are encoded in amino acid sequences. By focusing on protein structures without such nonideal features, we proposed principles for designing ideal protein structures stabilized by completely consistent local and nonlocal interactions (24), based on a set of rules relating local backbone structures to preferred tertiary motifs (7, 10). These design rules describe the relation of the lengths or torsion patterns of two secondary structure elements and the connecting loop to favorable packing geometries (*SI Appendix*, Fig. S1*A*). The design principles enable to encode strongly funneled energy landscapes into amino acid sequences, by the stabilization of folded structures (positive design) and by the destabilization of nonnative conformations (negative design) due to the restriction of folding conformational space by the rules (*SI Appendix*, Fig. S1*C*). In the design procedure, backbone structures for a target topology are generated based on a blueprint (*SI Appendix*, Fig. S1*B*), in which either the lengths or backbone torsion patterns of the secondary structures and loops are determined using the rules so that the tertiary motifs present in the target topology are favored, and then amino acid sequences stabilizing the generated backbone structures are designed. The designed amino acid sequences stabilize their folded structures both with nonlocal interactions such as hydrophobic core packing and with local interactions favoring the secondary structures and loops specified in the blueprint, which destabilize a myriad of nonnative topologies through local backbone strain captured by the rules, thereby resulting in funnel-shaped energy landscapes (*SI Appendix*, Fig. S1*C*). The principles have enabled the de novo design of ideal protein structures for various topologies with atomic-level accuracy (Fig. 1) (6, 7, 10, 13).

Interestingly, the de novo designs exhibit prominent characteristics in terms of thermal stability when compared with naturally occurring proteins. The circular dichroism (CD) measurements up to 170 °C conducted in this study revealed the melting temperature (T_m), which was above 100 °C for most of the designs (Fig. 1) (6, 7, 10). Therefore, the designs have great potential for use as scaffolds to engineer proteins with specific functions of interest. Indeed,

Significance

De novo designed proteins exhibit a remarkable property of extremely high thermal stability compared with naturally occurring proteins. The designed proteins are completely optimized for folding; the backbone structures are created by using a set of rules that relate local backbone structures to preferred tertiary motifs and the side chains are designed to favor both the local backbone structures and the entire tertiary structures. Here, we found that one of the de novo designed proteins, which was mutated to fill the core with mostly valine residues, still has the folding ability and shows high stability (T_m = 106 °C) even with its reduced and loosened core packing. This result supports the importance of local backbone structures to protein folding.

Author contributions: R.K., T.K., and N. Koga designed research; R.K., M.Y., T.K., N. Kobayashi, and T.S. performed research; R.K., N. Kobayashi, and T.S. analyzed data; and R.K., N. Kobayashi, T.S., T.F., and N. Koga wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: riekoga@ims.ac.jp or nkoga@ims.ac.jp.

[2]M.Y., T.K., and N. Kobayashi contributed equally to this work.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2002120117/-/DCSupplemental.

First published November 23, 2020.

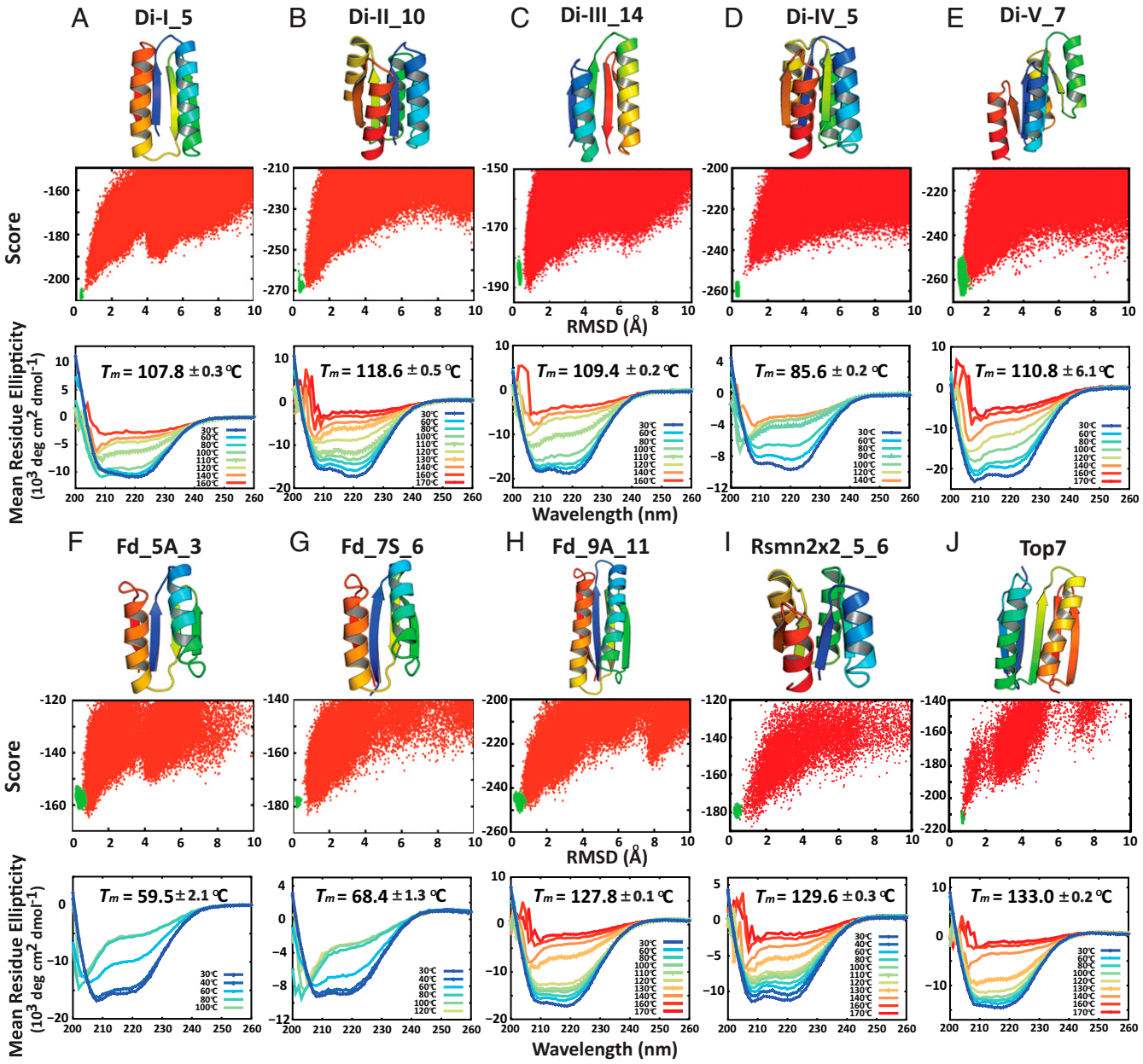


Fig. 1. In silico energy landscapes and far-UV circular dichroism (CD) spectra for 10 de novo designed ideal proteins. (*A–E*) Five designs by Koga et al. in 2012 (7). (*F–I*) Four designs by Lin et al. in 2015 (10). (*J*) Top7 by Kuhlman et al. in 2003 (6). (*Top*) Design models. (*Middle*) Energy landscapes obtained from Rosetta ab initio structure prediction simulations (41). Red points represent the lowest energy structures obtained in independent Monte Carlo structure prediction trajectories starting from an extended chain for each sequence; the *y* axis is the Rosetta all-atom energy; the *x* axis is the Cα root-mean-square deviation (RMSD) to the design model. Green points represent the lowest energy structures obtained in trajectories starting from the design model. (*Bottom*) The far-UV CD spectra during thermal denaturation with the melting temperature T_m, which is obtained by fitting to the denaturation curves shown in *SI Appendix*, Fig. S2.

miniprotein structures (~40 residues) designed de novo according to the rules were applied as scaffolds for creating protein binders specific for influenza hemagglutinin and botulinum neurotoxin, displaying high thermal stability (>70 °C) despite the small size (25).

The rules in the principles described above emphasize the importance of local backbone structures not the details of amino acid side chains to protein folding, which is also supported by studies using simple calculations with the hydrophobic-polar lattice model or the snake-cube model (26, 27). On the other hand, it is known that hydrophobic interactions are the dominant driving force for folding (28, 29) and the cores of naturally occurring proteins are tightly packed with hydrophobic amino acid residues (30, 31) like a jigsaw puzzle. Indeed, in our design principles, protein cores were designed to be tightly packed and as "fat" as possible with larger hydrophobic residues so that energy landscapes were sculpted to be deeply funneled into a target topology by lowering its energy (*SI Appendix*, Fig. S1*C*).

Which factor, the local backbone structures encoded by the rules or the tight core packing with fat hydrophobic residues, contributes more to the generation of funnels in the designs? Here, we studied the contribution of hydrophobic core packing

to folding ability and thermal stability by investigating the robustness of folding against the reduction of packing, using the design with the highest thermal stability among our nine de novo designs (Fig. 1, except Top7), Rsmn2x2_5_6 (10). We started to study single-residue mutants from Leu or Ile to Val that prune one carbon atom from the aliphatic side chain, which lose the tight packing like a jigsaw puzzle and decrease the hydrophobicity, and then, we combined the mutations. Consequently, we found that a mutant with 10 residue substitutions of Leu or Ile with Val still has the folding ability and high thermal stability despite its reduced and loosened hydrophobic core packing. This result suggests the importance of the local backbone structures for the folding ability and stability of the de novo designs.

Results

The de novo designed Rossmann2x2 fold protein, Rsmn2x2_5_6 (10), in which two helices sandwich a β-sheet, showed superhigh thermal stability with the melting temperature (T_m) of 129.6 °C (Fig. 1*I*). The design was computationally created using Rosetta by generating backbone structures based on a blueprint (*SI Appendix*, Fig. S1*B*), which was derived from the rules so that the Rossmann2x2 fold was selectively favored compared with other topologies, and then by designing sequences that stabilize the generated backbone structures with only hydrophobic residues in the core and only polar residues on the surface. The core of this protein consists of 20 valine residues, 7 leucine, 3 isoleucine, 2 alanine, 1 phenylalanine, and 1 tyrosine (Fig. 2). We substituted leucine and isoleucine with valine and measured the T_m of the mutants using CD.

Single-Residue Mutations from Leu or Ile to Val. The 10 single-residue mutants, L4V, L13V, L28V, I39V, I42V, L52V, L59V, L60V, I63V, and L80V (Fig. 3*A*), were generated (the residue numbers correspond to those of the NMR structure of Rsmn2x2_5_6; Protein Data Bank [PDB] ID: 2N3Z). The mutants were expressed in *Escherichia coli*, purified, and characterized by CD and size-exclusion chromatography combined with multiangle light scattering (SEC-MALS). All mutants were well expressed, soluble, and monomeric, and showed far-ultraviolet (UV) CD spectra characteristic of αβ-proteins as the original Rsmn2x2_5_6 design. Thermal denaturation experiments were carried out to identify the T_m of each mutant by using CD up to 170 °C under 1-MPa pressure, which made it possible to study

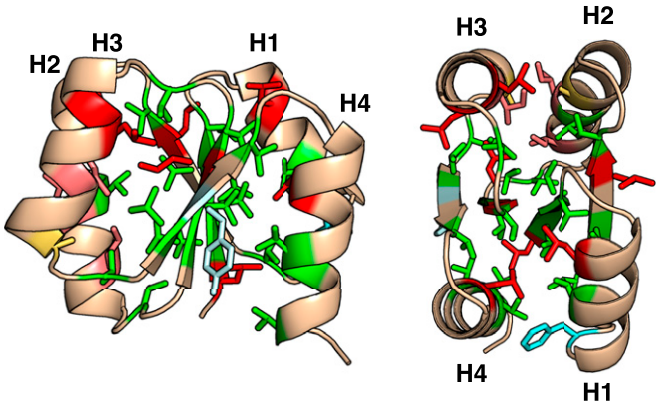
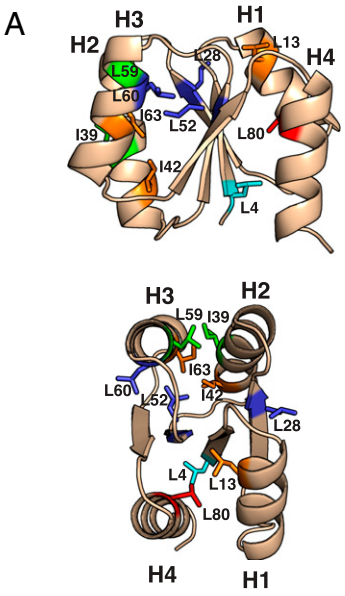


Fig. 2. Hydrophobic core of the de novo designed protein, Rsmn2x2_5_6. The core consists of 20 valine residues (green), 7 leucine (red), 3 isoleucine (salmon), 2 alanine (yellow), 1 phenylalanine (cyan), and 1 tyrosine (pale cyan). The structure on the *Left* is viewed from the side while the structure on the *Right* is viewed from the top following a 90° clockwise rotation of the structure on the *Left*. H1-4 represents the first to fourth helices. The first model of the NMR structure for the original design (PDB ID: 2N3Z) was used for the illustration.



	# SS contacts distant	close
L80	3	1
I63	1	1
I42	1	1
L13	1	1
L59	1	0
I39	1	0
L4	1	0
L52	0	1
L28	0	1
L60	0	2

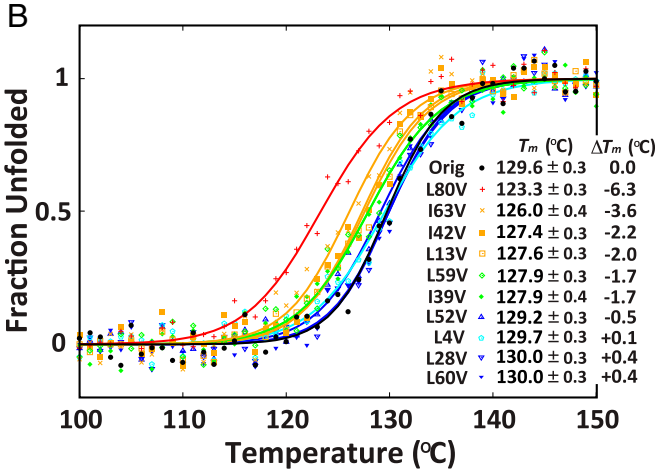


Fig. 3. Stability of single-residue substitutions of Leu or Ile with Val within the core. (*A*) Mutated residues (Leu, Ile) are shown in the two views, as in Fig. 2. The residue numbers correspond to those in the NMR structure of Rsmn2x2_5_6 (PDB ID: 2N3Z). Colors indicate the types of secondary structure (SS) contacts shown in the table: Residues with both contact types are shown in orange (L80, which forms the largest number of distant SS contacts, is shown in red), those that only form distant contacts are shown in green except for L4, which is shown in cyan, and those that only form close contacts are shown in blue (see main text for details). (*B*) Thermal denaturation measured by CD at 222 nm. The data were fit to the Gibbs–Helmholtz equation (solid lines) to obtain the melting temperature T_m, the enthalpy change ΔH, and the heat capacity change ΔC_p. The color scheme is the same as that used in *A*. The original (Orig) is shown in black.

the contribution of hydrophobic core packing to the stability in the superstable de novo design. The L80V mutant, represented by red crosses in Fig. 3*B*, had the lowest T_m, 6.3 °C lower than the original protein (Orig), represented by black filled circles. (This mutant also showed lower cooperativity.) This was followed by the I63V, I42V, L13V, L59V, and I39V mutants, in ascending order of stability. Four mutants, L52V, L4V, L28V, and L60V, did not show a significant change in T_m (<1 °C) (Fig. 3*B* and *SI Appendix*, Table S1). (Some single-residue mutations substituting Leu on helix with Ala were also carried out. See *SI Appendix*, Table S2.)

To understand why each mutant made a different contribution to the stability, we attempted to predict the change in stability by computational modeling using Rosetta. We carried out Rosetta $\Delta\Delta G$ calculations, which were developed for evaluating the free energy difference between the wild type and its mutant (32) (see *SI Appendix*, SI Method 1 for details). In addition, we assessed the packing quality for each mutant using RosettaHoles (33). As expected, all 10 single mutants showed decreased stability and packing quality. However, the correlations of the decreases of the melting temperature ΔT_m with the calculated $\Delta\Delta G$ values and the packing quality were not found (*SI Appendix*, Figs. S3*A* and S4*A*). These results are probably due to the insufficient accuracy of the energy function to evaluate an energy at the single-residue resolution (34) or due to the difficulty of predicting structural rearrangements induced by a single-residue mutation (35) (see strong correlations when the combined mutants using the single mutations are included for the comparison [*SI Appendix*, Figs. S3*B* and S4*B*]). Here, we sought to understand the different contribution made by each mutated residue to the stability by considering the number of secondary structure elements that the residue makes contacts with and classifying those into the distant (not adjacent to the secondary structure element containing the mutated residue) and close (adjacent) ones (table in Fig. 3*A* and *SI Appendix*, Fig. S5 and Table S3; contacts are defined in *Materials and Methods*). Interestingly, we found that the hydrophobic residues that have multiple secondary structure contacts including both the contact types, "hub-residues," L80, I63, I42, and L13 (red and orange in Fig. 3), contributed strongly to stability. L80 (red), in particular, is expected to contribute to stability the most because it forms the greatest number of distant secondary structure contacts, which agrees with the largest change of T_m upon the mutation. Residues that form only distant secondary structure contacts, such as L59 and I39 (green in Fig. 3), but not L4 (cyan), appeared to make relatively lower yet still significant contributions to the stability. In contrast, those residues that form only close secondary structure contacts (blue in Fig. 3) did not appear to contribute to stability, possibly indicating the importance of distant secondary structure contacts for stability. Interestingly, the L4V mutant did not change T_m significantly despite forming distant secondary structure contacts. L4 at the N terminus interacts with residues at the C-terminal helix (H4); therefore, the enthalpy loss upon the mutation may be compensated by the entropy gain due to the flexibility of both termini.

Combined Mutations. We next constructed the combined variants using the single mutations conducted above. The Rossmann2x2 fold has two hydrophobic packing sites separated by a β-sheet: the helix (H) 2,3 side and the H1,4 side (Fig. 3*A*). The H2,3 side has six residues that are Leu or Ile: I39, I42, L52, L59, L60, and I63, while the H1,4 side has four residues: L4, L13, L28, and L80. We generated two combined variants consisting of all of the single mutations on each side, which were subsequently expressed and purified. Remarkably, the variants were still soluble and monomeric, and exhibited CD spectra characteristic of αβ-proteins. The T_m of the H2,3 and H1,4 side combined variants were 115.3 and 123.2 °C, respectively (*SI Appendix*, Table S4 and Fig. S6). The thermal denaturation of the H1,4 side was considerably less cooperative than that of the original (*SI Appendix*, Fig. S6), which may be due to the decreased cooperativity of the L80V mutant described above.

The Mutant with Most of the Core Filled with Valine. We finally constructed a mutant with total 10 substitutions of Leu or Ile with Val, resulting in 30 of the 34 hydrophobic core residues being valine (88% valine in the core, thus referred to as VAL88). The VAL88 protein was well expressed and still soluble (*SI Appendix*, Fig. S7*A*), although it starts to aggregate after being left at ∼500 μM concentration for 1 wk at room temperature. Furthermore, VAL88 in the soluble fraction was found to be monomeric by SEC-MALS (*SI Appendix*, Fig. S7*B*) and showed a CD spectrum characteristic of αβ-proteins (Fig. 4*B*) and high stability with T_m of 106.1 °C (*SI Appendix*, Figs. S7*C* and S8 and Table S5). The 23.6 °C decrease in T_m compared with the original is almost equal to the sum of the decreases for the combined variants on each side (20.8 °C) (*SI Appendix*, Table S4), suggesting that the two sides separated by the β-sheet contribute almost independently to stability. As well as the T_m, the cooperativity of the thermal denaturation of VAL88 also decreased compared with the original (*SI Appendix*, Fig. S8). The experimentally estimated $\Delta\Delta G$, the change in free energy of unfolding for VAL88 compared with the original, at the T_m of the original, was −5.5 kcal/mol (*SI Appendix*, SI Method 2 and Fig. S9), which is close to the calculated hydrophobic free energy change (−5.8 kcal/mol) based on buried surface area during folding (*SI Appendix*, SI Method 2) (36, 37). This result is consistent with those in previous studies (36, 37) that demonstrated how much protein stability is lost when carbon atoms in hydrophobic residues are removed from the core, indicating that total reduction in stability caused by the 10 substitutions can be explained by the reduction in hydrophobic interactions.

The two-dimensional (2D) ^{1}H–^{15}N heteronuclear single-quantum coherence (HSQC) spectrum of VAL88 showed the expected number of well-dispersed sharp peaks (Fig. 4*C*), indicating that the protein with most of the core mutated to valine is not in a molten globule state (38). The solution structure was then determined with high quality by NMR spectroscopy using MagRO-NMRViewJ (39, 40), and was found to have the same topology as the original (see Fig. 4*D*, *SI Appendix*, SI Discussion 1 and 2, Figs. S10–S12, and Table S6, and *Materials and Methods* for details). The hydrophobic core consisting of many valine residues (green sticks in Fig. 4*E*) has cavities (*Right* in Fig. 4*F*) compared with the original (*Left* in Fig. 4*F*). Based on hydrogen/deuterium exchange (HDX) experiments, core residues in VAL88 were found to be significantly more unprotected than the original (Fig. 4*G* and *SI Appendix*, Fig. S13 and Tables S7 and S8). The packing quality evaluated by RosettaHoles (33) also suggested the loosened packing: The score of VAL88, 6.0, is higher than those of the original, −0.45, and the other de novo designs (lower values represent better packing) (*SI Appendix*, Fig. S14). Despite the loosened packing, the structure maintained the same topology and backbone structure as the original with an averaged root-mean-square deviation of Cα atoms of 1.4 Å (Fig. 4*D* and *SI Appendix*, Fig. S15).

To identify how the energy landscape of VAL88 is sculpted, we performed 10,000 independent Rosetta ab initio structure prediction simulations (41) starting from an extended chain. The energy landscape still remained funnel-shaped toward the same folded structure as the original, although the funnel got wider and the energy gap between the native and the denatured states got smaller than the original, likely because the hydrophobic interactions were weakened by the mutations to smaller hydrophobic residues (Fig. 4*A* and the energy landscape of the original, Rsmn2x2_5_6, in Fig. 1*I*). This result agrees with the fact that VAL88 can fold into the same structure as the original protein, even though the cooperativity and melting temperature decreased.

We also constructed a mutant with an additional substitution F21V (11 residues were substituted with Val in total). Notably, the ^{1}H–^{15}N HSQC spectrum still showed well-dispersed sharp peaks, and the thermal denaturation showed the T_m of 101.5 °C, indicating that the protein is still folded with high stability, although the soluble monomeric fraction was quite low (*SI Appendix*, Fig. S16).

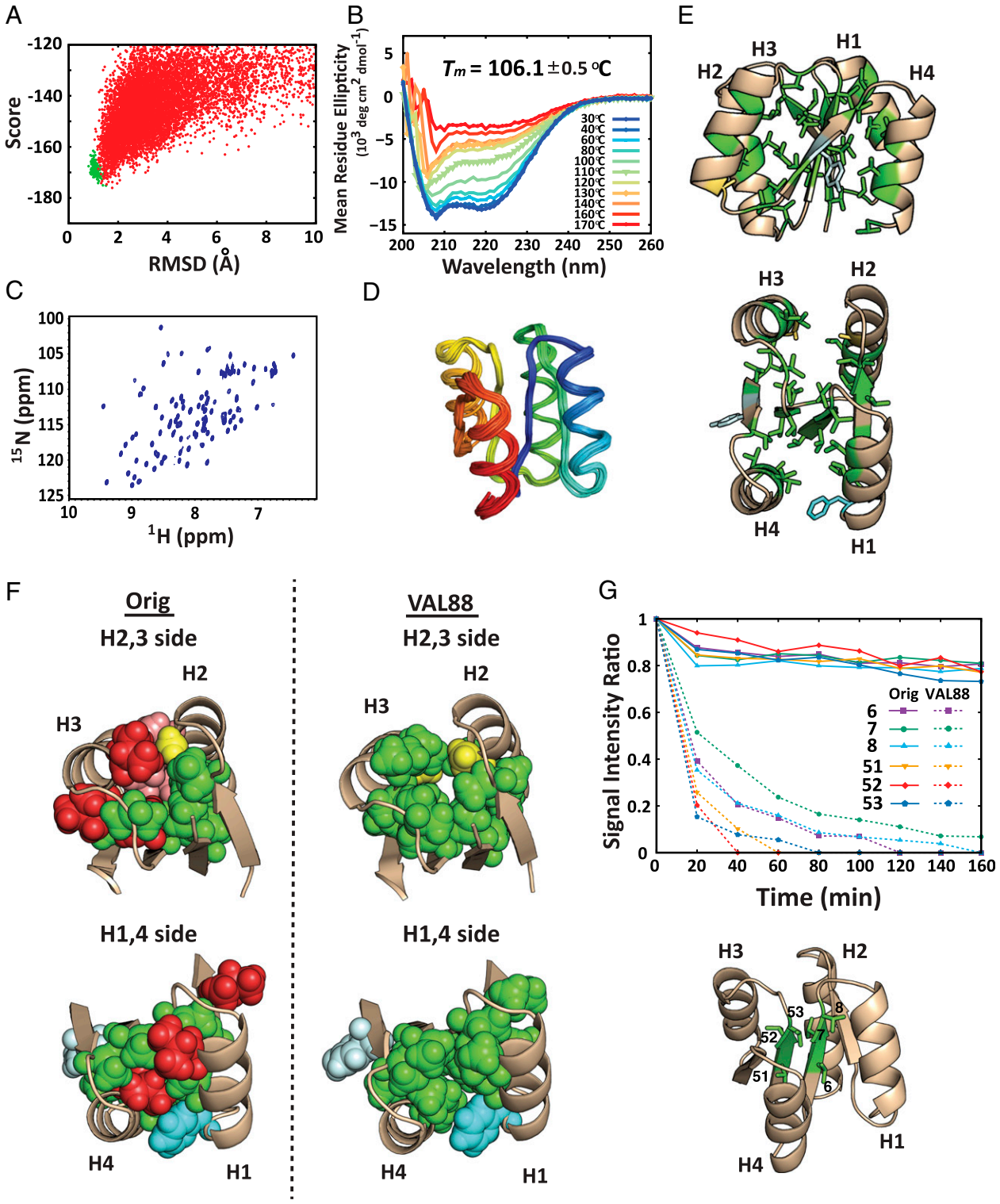


Fig. 4. Characterization of the designed protein with most of the core mutated to Val, VAL88. (*A*) The energy landscape obtained from Rosetta ab initio structure prediction simulations. As in Fig. 1, red points represent simulated structures starting from an extended chain. The Cα RMSD of simulated structures was calculated against the first model of the NMR structure for the original (PDB ID: 2N3Z). Green points represent simulated structures starting from the VAL88 structure generated from the first NMR model using Foldit (55). (*B*) The far-UV CD spectra at various temperatures. (*C*) Two-dimensional ^{1}H–^{15}N HSQC spectrum at 25 °C and 600 MHz. (*D*) NMR structure (PDB ID: 6LLQ). (*E*) Hydrophobic core side chains are shown in stick. The residue color and the views are the same as those in Fig. 2. Residues colored in green are valine. (*F*) Core packing in the H1,4 and H2,3 sides of the original and VAL88 NMR structures are shown in the same view as the *Bottom* figure in *E*. Hydrophobic residues are represented by space-filling with the same residue color as in Fig. 2. The first model of each NMR structure was used for illustration. (*G*) HDX rates for core residues. (*Top*) Changes in signal intensity ratio (signal intensity at the beginning is set to 1) of ^{1}H–^{15}N correlation NMR signals for six residues in the core are shown for the original (Orig) as solid lines and for VAL88 as dashed lines. (*Bottom*) The six residues shown by green sticks. Residue 52 is Leu in the original and Val in VAL88; the other five residues are Val in the both. The first model of the NMR structure for VAL88 was used for illustration in *E* and *G*.

The Mutant with Altered Loop Lengths. Finally, we directly investigated the importance of local backbone structures by changing the lengths of the loops that connect a β-strand to an α-helix (*SI Appendix*, Fig. S1 *A* and *B*). We created a mutant that violates the βα-rules (*SI Appendix*, Fig. S17*A*) by changing the two-residue loops to three-residue ones with the insertion of an Ala into the middle, and the three-residue loops to two-resiude ones with the deletion of the middle residue. We found that most of the soluble fraction of the mutant was in an aggregated form and that the monomeric fraction was quite low by SEC-MALS (*SI Appendix*, Fig. S17 *B*, *Left*). Furthermore, the ^{1}H–^{15}N HSQC spectrum showed a small number of nondispersed peaks (*SI Appendix*, Fig. S17 *B*, *Right*), indicating that the mutant can no longer fold. This result does not, however, imply that de novo designed structures are vulnerable to any loop changes. A wide range of loop types should be tested for the purpose of creating functional proteins.

Discussion

Various ideal protein structures stabilized by completely consistent local and nonlocal interactions have been designed using the design rules (6, 7, 10, 13). Among the consistent local and nonlocal interactions, we investigated the contribution of a nonlocal interaction, hydrophobic core packing, to the folding ability and stability of de novo designed proteins by reducing hydrophobic core packing through extensive mutation of large hydrophobic residues (Leu, Ile) to smaller ones (Val). The results obtained from single-residue mutations suggested that "hub-residues," the residues having contacts with both distant and close secondary structures, make a large contribution to the stability. It would be interesting to test the generality of this idea for other proteins. Furthermore, by accumulating the single mutations, we demonstrated that a designed protein with most of the core mutated to valine can still fold into the same topology as the original and remains folded above 100 °C even with the stability reduced by 5.5 kcal/mol. This result indicates that hydrophobic tight core packing may not be quite important for designed proteins: The folding ability and extremely high stability may arise from the restriction of conformational space to be searched during folding by the local backbone structures. This can lead to an exceptionally stable protein even in the absence of precise core packing. On the other hand, the contribution of local interactions in naturally occurring proteins may not be significant compared with designed proteins. This is because naturally occurring protein structures are prevented from completely being optimized for folding and stability due to the requirements to be functional or the susceptibility to neutral evolution, although the design rules are also observed in naturally occurring protein structures (7). Indeed, to the best of our knowledge, there are no natural proteins that can still fold properly when most of the hydrophobic residues in the core are mutated into valine. In addition, proteins that contain a high ratio of valine in the core are not observed in naturally occurring globular proteins including Rossmann folds (*SI Appendix*, Fig. S18). Finally, we should note studies in which most of hydrophobic core residues of naturally occurring proteins were substituted with either methionine or isoleucine (42, 43). However, these studies aimed to create alternative core packing with methionine or isoleucine having higher hydrophobicity and higher number of side-chain conformations than valine; VAL88 created in this study is uniquely characterized by the reduced and loosened tight core packing with valine.

Unlike de novo designed proteins, most naturally occurring proteins are not highly stable and unfold at less than 100 °C. Thermostabilization of natural proteins has been a long-standing challenge for industrial applications. Many studies have attempted to increase the thermal stability of proteins by changing side chains (amino acid residues) without modifying backbone conformations (44). Remodeling backbone structures with ideal ones would be an alternative way to achieve thermal stabilization of naturally occurring proteins (45, 46).

Materials and Methods

Definition of Secondary Structure Contacts. Secondary structure contacts that each mutated residue forms were defined as follows. First, using all 20 models of the NMR structure for the original design, Rsmn2x2_5_6 (PDB ID: 2N3Z), residue-pair contacts were calculated if any heavy atoms of two residues separated by at least five residues along the sequence are within 4.5 Å in more than half of the 20 NMR models. All residue-pair contacts are shown in *SI Appendix*, Table S3. Subsequently, secondary structure elements that each mutated residue makes contact with were identified by looking up the calculated residue-pair contacts, and then classified into the distant (not adjacent to the secondary structure element containing the mutated residue) and close (adjacent) ones (*SI Appendix*, Fig. S5). Here, the residue pairs between β-strands were excluded for the calculation, because nonlocal interactions between the strand residues are provided by backbone–backbone hydrogen bonding in addition to hydrophobic contacts between side chains; therefore, the mutations of hydrophobic residues may not largely contribute to stability. Indeed, the mutations of the strand residues forming contacts with the neighboring β-strands (L52V, L4V, and L28V) did not show a significant stability change (*SI Appendix*, Table S1).

Calculation of Energy Landscapes. Energy landscapes of the original (Rsmn2x2_5_6, Fig. 1*I*), VAL88 (Fig. 4*A*), and Top7 (Fig. 1*J*) were obtained from Rosetta ab initio structure prediction simulations (41).

We performed 10,000 independent Monte Carlo structure prediction simulations starting from an extended chain for each sequence using energy function talaris2014 on Research Center for Computational Science (RCCS). The command line used for executing Rosetta ab initio structure prediction simulations starting from an extended chain is the following:

```
$ROSETTA/source/bin/AbinitioRelax.linuxgccrelease -database $ROSETTADB
-in:file:native [input.pdb] -in:file:fasta [input.fasta]
-in:file:frag3 [input.frag3] -in:file:frag9 [input.frag9] -abinitio:fastrelax
-out:file:silent [output.silent] -nstruct XX -ex1 -ex2aro -use_filters false
-abinitio::increase_cycles 10 -abinitio::rg_reweight 0.5 -abinitio::rsd_wt_helix 0.5
abinitio::rsd_wt_loop 0.5
-relax::fast -relax::default_repeats 15 -score:weights talaris2014,
```

where [input.pdb] is a pdb file of a design model, [input.fasta] is the designed sequence, [input.frag3] and [input.frag9] are three-residue or nine-residue fragment files generated from the designed sequence, and XX is a number of folding simulation trajectories. The relax simulations starting from the design model are carried out using the command line below:

```
$ROSETTA/source/bin/relax.linuxgccrelease -database $ROSETTADB
-s [input.pdb]
-in:file:native [input.pdb]
-out:file:silent [output.silent] -nstruct XX -ex1 -ex2aro -use_filters false
-relax::fast -relax::default_repeats 15 -score:weights talaris2014.
```

Plasmid Construction, Expression, and Purification of Mutants. The gene encoding the originally designed sequence, Rsmn2x2_5_6, cloned into plasmid pET21b, was obtained from the previous study (10). All single-residue mutations were introduced by the site-directed mutagenesis, in which a pair of complementary mutagenic primers are used to amplify the whole plasmid using Phusion high-fidelity DNA polymerase (New England Biolabs) or PrimeSTAR mutagenesis basal kit (Takara Bio). The combined mutations were introduced by QuikChange multi site-directed mutagenesis kit (Agilent Technologies). Oligos were obtained from FASMAC or Eurofins Genomics. The mutated DNA sequences were confirmed by DNA sequencing by FASMAC.

The mutated proteins except VAL88 were expressed in *E. coli* BL21 Star (DE3) cells (Invitrogen) as nonlabeled proteins, which were expressed using autoinduction media (47). The expressed proteins with a 6×His tag at the C terminus were purified through a nickel affinity column. The purified proteins were then dialyzed against PBS buffer, 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, at pH 7.4; this buffer was used for all of

the experiments except VAL88. The expression, solubility, and purity of the designed proteins were assessed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS/PAGE) and mass spectrometry (Thermo Scientific Orbitrap Elite).

CD. All CD data were collected on a JASCO J-1500 KS CD spectrometer under 1-MPa pressure, which makes the thermal denaturation possible up to 170 °C. Thermal denaturation was performed from 30 to 170 °C with an increase of 1 °C per min using ∼20 μM protein samples in PBS buffer (pH 7.4) in a 1-mm path length cuvette. During the denaturation, the ellipticity at 222 nm was monitored and far-UV CD spectra were measured from 260 to 200 nm at various temperatures of 30, 40, 60, 80, 100, 110, 120, 130, 140, 160, and 170 °C. The thermal denaturation curves were fit to the Gibbs–Helmholtz equation:

$$\Delta G = \Delta H_m\left(1-\frac{T}{T_m}\right) + \Delta C_p\left\{(T-T_m) - T\ln\left(\frac{T}{T_m}\right)\right\},$$

to obtain the melting temperature T_m, in which the number of folded proteins is equal to the number of unfolded proteins, the enthalpy change ΔH_m and the heat capacity change ΔC_p. Since the heat capacity is a function of pressure in general, we investigated the dependency by performing thermal denaturation under the two different pressures, the atmospheric pressure, 0.1 MPa, and 1 MPa for heating up to 170 °C, using one of our de novo designs, Fd_5A_3. The heat capacities obtained by fitting to the Gibbs–Helmholtz equation were 0.79 kcal/mol/K under 0.1-MPa pressure and 0.68 kcal/mol/K under 1-MPa pressure; these values agree well.

SEC-MALS. SEC-MALS experiments were performed using a miniDAWN TREOS static light scattering detector (Wyatt Technology Corporation) combined with a HPLC system (1260 Infinity LC; Agilent Technologies). The volume 100 μL of 300 to 500 μM protein samples in PBS buffer (pH 7.4) after Ni purification except VAL88 protein samples for NMR structure determination was injected into a Superdex 75 increase 10/300 GL column (GE Healthcare) or a Shodex KW-802.5 (Showa Denko K.K.) equilibrated with PBS buffer at a flow rate of 0.5 mL/min. As for VAL88 proteins purified by gel filtration chromatography, the volume 50 μL of 500 μM samples in PBS buffer (pH 6.0) was injected into a Superdex 75 increase 10/300 GL column. The protein concentrations were calculated from the absorbance at 280 nm detected by the HPLC system. Static light scattering data were collected at three different angles, 43.6°, 90.0°, and 136.4°, at 659 nm. These data were analyzed by the ASTRA software (version 6.1.2; Wyatt Technology Corporation) with a change in the refractive index with concentration, a *dn/dc* value, 0.185 mL/g.

Two-Dimensional ^{1}H–^{15}N HSQC Measurement. Two-dimensional ^{1}H–^{15}N HSQC spectrum was measured to verify whether the core of VAL88 proteins is well packed, i.e., not in a molten globule state. The VAL88 was expressed in *E. coli* BL21 Star (DE3) cell as uniformly (U-)^{15}N-labeled proteins using MJ9 minimal media (48), which contain ^{15}N ammonium sulfate as a sole nitrogen source and ^{12}C glucose as a sole carbon source, respectively. The expressed proteins with a 6×His tag at the C terminus were purified through a nickel affinity column and then purified by gel filtration chromatography on an ÄKTA Pure 25 FPLC (GE Healthcare) using a Superdex 75 increase 10/300 GL column (GE Healthcare). The expression, solubility, and purity of VAL88 proteins were assessed by SDS/PAGE and mass spectrometry (Thermo Scientific Orbitrap Elite). The HSQC spectrum was collected for 500 μM protein sample in 90% H_2O/10% D_2O buffer containing 100 mM NaCl, 1.2 mM Na_2HPO_4, and 7.4 mM KH_2PO_4 at pH 6.0, at 25 °C on a JEOL JNM-ECA 600-MHz spectrometer.

Determination of Solution Structure by NMR.

Sample preparation. For the NMR structure determination, the uniformly isotope labeled [U-^{15}N,U-^{13}C]-VAL88 was expressed using the same method as described above except ^{13}C glucose as a sole carbon source. The U-^{15}N,U-^{13}C-enriched purified protein samples dissolved in 95% H_2O/5% D_2O buffer containing 100 mM NaCl, 1.2 mM Na_2HPO_4, and 7.4 mM KH_2PO_4 at pH 6.0 were transferred into Shigemi micro NMR tubes (protein concentration, ∼500 μM). For residual dipolar coupling (RDC) experiments, normal NMR tubes were used (protein concentration, ∼200 μM).

NMR measurements. NMR measurements were performed on Bruker AVANCE III NMR spectrometers equipped with QCI cryo-Probe (^{1}H/^{13}C/^{15}N/^{31}P) at 303 K. The spectrometers with 600- and 800-MHz magnets were used for the signal assignments and nuclear Overhauser effect (NOE)-related measurements, while a 950-MHz one was used for RDC experiments. For the signal assignments, 2D ^{1}H–^{15}N HSQC (echo/anti-echo), ^{1}H–^{13}C constant-time HSQC for aliphatic and aromatic signals, 3D BEST-HNCO, 3D CBCA(CO)NH, 3D HNCACB, 3D H(CCCO)NH, and 3D CC(CO)NH spectra were measured. For structure determination, 3D ^{15}N-edited nuclear Overhauser effect spectroscopy (NOESY), 3D ^{13}C-edited NOESY for aliphatic and aromatic signals (mixing time, 100 ms) were performed. For RDC experiments, 2D IPAP ^{1}H–^{15}N HSQC using water-gate pulses for water suppression was measured with or without 7 mg/mL Pf1 phage (ASLA Biotech Ltd.). For confirming the positions of ^{1}H–^{15}N signals in the 2D HSQC, 3D HNCO in the presence of Pf1 phage at the same buffer condition were used. The signals for α- and β-states split by ^{1}H–^{15}N ^{1}J-coupling were separately obtained for the protein in the isotropic and weakly aligned states. The RDC values were estimated by simple subtraction of the shifted values between isotropic and weakly aligned states, and then divided by the static magnetic field to get the value in hertz.

NMR data process. All of the acquired NMR free induction decay data were processed with NMRPipe (49). For the NMR analysis, an integrated package of NMR tools named MagRO-NMRViewJ, version 2.01.12 [the upgraded version of Kujira (39)], on NMRView (50) was used for automated peak identification and noise filtration using convolutional neural networks [CNN-filter (40)]. The filtered peak lists were applied to calculations for automated signal assignments by FLYA (51) and then the signal assignments were used for prediction of dihedral angles by TALOS+ (52), and automated NOE assignments and structure calculation by CYANA (53).

HDX NMR Experiments. Uniformly ^{15}N-labeled original and VAL88 proteins were lyophilized after purification. As reference experiments, the lyophilized samples were dissolved in a 95% H_2O/5% D_2O buffer solution containing 100 mM NaCl, 1.2 mM Na_2HPO_4 and 7.4 mM KH_2PO_4 at pH 6.0, and 2D ^{1}H–^{15}N HSQC NMR spectra were measured by Bruker AVANCE III 800-MHz NMR spectrometer equipped with a TXI cryogenic probe at 303 K. In HDX experiments, the lyophilized sample was dissolved in a prechilled 100% D_2O buffer solution, of which composition was same as that of the reference experiments. Upon addition of the 100% D_2O buffer, the sample was immediately transferred into a Shigemi tube (5-mm diameter) and placed in the Bruker NMR magnet of 800 MHz, and 2D ^{1}H–^{15}N HSQC NMR spectral data were collected. The time from the dissolution of the lyophilized sample to the beginning of NMR data collection was 20 min, and the eight of the ^{1}H–^{15}N HSQC NMR spectra (measurement time of one NMR spectrum was 20 min) were measured continuously. Those NMR spectral data were processed by using the program NMRPipe (49), and H/D exchanging time-dependent changes of signal intensity of those spectra were analyzed by using the program NMRFAM-SPARKY (54).

Data Availability. The solution NMR structure of R2x2_VAL88 has been deposited in the Protein Data Bank (accession code 6LLQ). The NMR data have been deposited in the Biological Magnetic Resonance Data Bank (accession no. 36305). The plasmid of R2x2_VAL88 is available through Addgene (160761). All study data are included in the article and *SI Appendix*.

ACKNOWLEDGMENTS. We thank Functional Genomics Facility, National Institute for Basic Biology Core Research Facilities, especially Yumiko Makino, for mass spectrometry analysis, and the Instrument Center, Okazaki, Japan, especially Michiko Nakano and Haruyo Nagao, for HSQC spectra measurements. We also thank Ingemar André and Shoji Takada for helpful comments on the manuscript. The computations were performed using RCCS, Okazaki, Japan. NMR structure determination was supported by Basis for Supporting Innovative Drug Discovery and Life Science Research from Japan Agency for Medical Research and Development under Grant JP19am0101072. This work was supported by the Japan Society for the Promotion of Science KAKENHI Grants-in-Aid for Scientific Research 15H05592 to N. Koga, 18H05420 to T.K. and N. Koga, and 18K06152 to N. Kobayashi, and the Japan Science and Technology Agency Precursory Research for Embryonic Science and Technology (Grant JPMJPR13AD to N. Koga).

1. S. P. Ho, W. F. Degrado, Design of a 4-helix bundle protein—synthesis of peptides which self-associate into a helical protein. *J. Am. Chem. Soc.* **109**, 6751–6758 (1987).
2. M. H. Hecht, J. S. Richardson, D. C. Richardson, R. C. Ogden, De novo design, expression, and characterization of felix: A four-helix bundle protein of native-like sequence. *Science* **249**, 884–891 (1990).
3. P. B. Harbury, J. J. Plecs, B. Tidor, T. Alber, P. S. Kim, High-resolution protein design with backbone freedom. *Science* **282**, 1462–1467 (1998).
4. R. B. Hill, D. P. Raleigh, A. Lombardi, W. F. DeGrado, De novo design of helical bundles as models for understanding protein folding and function. *Acc. Chem. Res.* **33**, 745–754 (2000).

5. Y. Wei, S. Kim, D. Fela, J. Baum, M. H. Hecht, Solution structure of a de novo protein from a designed combinatorial library. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13270–13273 (2003).
6. B. Kuhlman *et al*., Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).
7. N. Koga *et al*., Principles for designing ideal protein structures. *Nature* **491**, 222–227 (2012).
8. P. S. Huang *et al*., High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481–485 (2014).
9. A. R. Thomson *et al*., Computational design of water-soluble α-helical barrels. *Science* **346**, 485–488 (2014).
10. Y. R. Lin *et al*., Control over overall shape and size in de novo designed proteins. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E5478–E5485 (2015).
11. T. J. Brunette *et al*., Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).
12. L. Doyle *et al*., Rational design of α-helical tandem repeat proteins with closed architectures. *Nature* **528**, 585–588 (2015).
13. P. S. Huang *et al*., De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat. Chem. Biol.* **12**, 29–34 (2016).
14. T. M. Jacobs *et al*., Design of structurally distinct proteins using strategies inspired by evolution. *Science* **352**, 687–690 (2016).
15. E. Marcos *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).
16. J. Dou *et al*., De novo design of a fluorescence-activating β-barrel. *Nature* **561**, 485–491 (2018).
17. E. Marcos, D. A. Silva, Essentials of de novo protein design: Methods and applications. *Wires Comput. Mol. Sci.* **8**, e1374 (2018).
18. R. Koga, N. Koga, Consistency principle for protein design. *Biophys. Physicobiol.* **16**, 304–309 (2019).
19. J. D. Bryngelson, P. G. Wolynes, Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7524–7528 (1987).
20. P. E. Leopold, M. Montal, J. N. Onuchic, Protein folding funnels: A kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8721–8725 (1992).
21. J. N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten, N. D. Socci, Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3626–3630 (1995).
22. K. A. Dill, H. S. Chan, From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19 (1997).
23. J. N. Onuchic, Z. Luthey-Schulten, P. G. Wolynes, Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.* **48**, 545–600 (1997).
24. N. Go, Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.* **12**, 183–210 (1983).
25. A. Chevalier *et al*., Massively parallel de novo protein design for targeted therapeutics. *Nature* **550**, 74–79 (2017).
26. G. Chikenji, Y. Fujitsuka, S. Takada, Shaping up the protein folding funnel by local interaction: Lesson from a structure prediction study. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3141–3146 (2006).
27. N. Go, Snake cube puzzle and protein folding. *Biophys. Physicobiol.* **16**, 256–263 (2019).
28. K. A. Dill, Dominant forces in protein folding. *Biochemistry* **29**, 7133–7155 (1990).
29. K. A. Dill *et al*., Principles of protein folding—a perspective from simple exact models. *Protein Sci.* **4**, 561–602 (1995).
30. F. M. Richards, Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* **6**, 151–176 (1977).
31. W. Sheffler, D. Baker, RosettaHoles: Rapid assessment of protein core packing for structure prediction, refinement, design, and validation. *Protein Sci.* **18**, 229–239 (2009).
32. H. Park *et al*., Simultaneous optimization of biomolecular energy functions on features from small molecules and macromolecules. *J. Chem. Theory Comput.* **12**, 6201–6212 (2016).
33. W. Sheffler, D. Baker, RosettaHoles2: A volumetric packing measure for protein structure refinement and validation. *Protein Sci.* **19**, 1991–1995 (2010).
34. D. Baker, What has de novo protein design taught us about protein folding and biophysics? *Protein Sci.* **28**, 678–683 (2019).
35. A. Nisthal, C. Y. Wang, M. L. Ary, S. L. Mayo, Protein stability engineering insights revealed by domain-wide comprehensive mutagenesis. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 16367–16377 (2019).
36. M. Matsumura, W. J. Becktel, B. W. Matthews, Hydrophobic stabilization in T4 lysozyme determined directly by multiple substitutions of Ile 3. *Nature* **334**, 406–410 (1988).
37. M. Matsumura, J. A. Wozniak, D. P. Sun, B. W. Matthews, Structural studies of mutants of T4 lysozyme that alter hydrophobic stabilization. *J. Biol. Chem.* **264**, 16059–16066 (1989).
38. M. Ohgushi, A. Wada, "Molten-globule state": A compact form of globular proteins with mobile side-chains. *FEBS Lett.* **164**, 21–24 (1983).
39. N. Kobayashi *et al*., KUJIRA, a package of integrated modules for systematic and interactive analysis of NMR data directed to high-throughput NMR structure studies. *J. Biomol. NMR* **39**, 31–52 (2007).
40. N. Kobayashi *et al*., Noise peak filtering in multi-dimensional NMR spectra using convolutional neural networks. *Bioinformatics* **34**, 4300–4301 (2018).
41. C. A. Rohl, C. E. Strauss, K. M. Misura, D. Baker, Protein structure prediction using Rosetta. *Methods Enzymol.* **383**, 66–93 (2004).
42. N. C. Gassner, W. A. Baase, B. W. Matthews, A test of the "jigsaw puzzle" model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12155–12158 (1996).
43. D. S. Riddle *et al*., Functional rapidly folding proteins from simplified amino acid sequences. *Nat. Struct. Biol.* **4**, 805–809 (1997).
44. A. Goldenzweig, S. J. Fleishman, Principles of protein stability and their application in computational design. *Annu. Rev. Biochem.* **87**, 105–129 (2018).
45. D. A. Silva *et al*., De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **565**, 186–191 (2019).
46. A. Ruggiero, G. Smaldone, L. Esposito, N. Balasco, L. Vitagliano, Loop size optimization induces a strong thermal stabilization of the thioredoxin fold. *FEBS J.* **286**, 1752–1764 (2019).
47. F. W. Studier, Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
48. M. Jansson *et al*., High-level production of uniformly ^{15}N- and ^{13}C-enriched fusion proteins in Escherichia coli. *J. Biomol. NMR* **7**, 131–141 (1996).
49. F. Delaglio *et al*., NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
50. B. A. Johnson, R. A. Blevins, NMR view: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).
51. E. Schmidt, P. Güntert, A new algorithm for reliable and general NMR resonance assignment. *J. Am. Chem. Soc.* **134**, 12817–12829 (2012).
52. Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 (2009).
53. P. Güntert, L. Buchner, Combined automated NOE assignment and structure calculation with CYANA. *J. Biomol. NMR* **62**, 453–471 (2015).
54. W. Lee, M. Tonelli, J. L. Markley, NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).
55. S. Cooper *et al*., Predicting protein structures with a multiplayer online game. *Nature* **466**, 756–760 (2010).