SUPPLEMENTARY INFORMATION

doi:10.1038/nature11600

Contents

Supplementary Discussion 1 | Emergent rules.

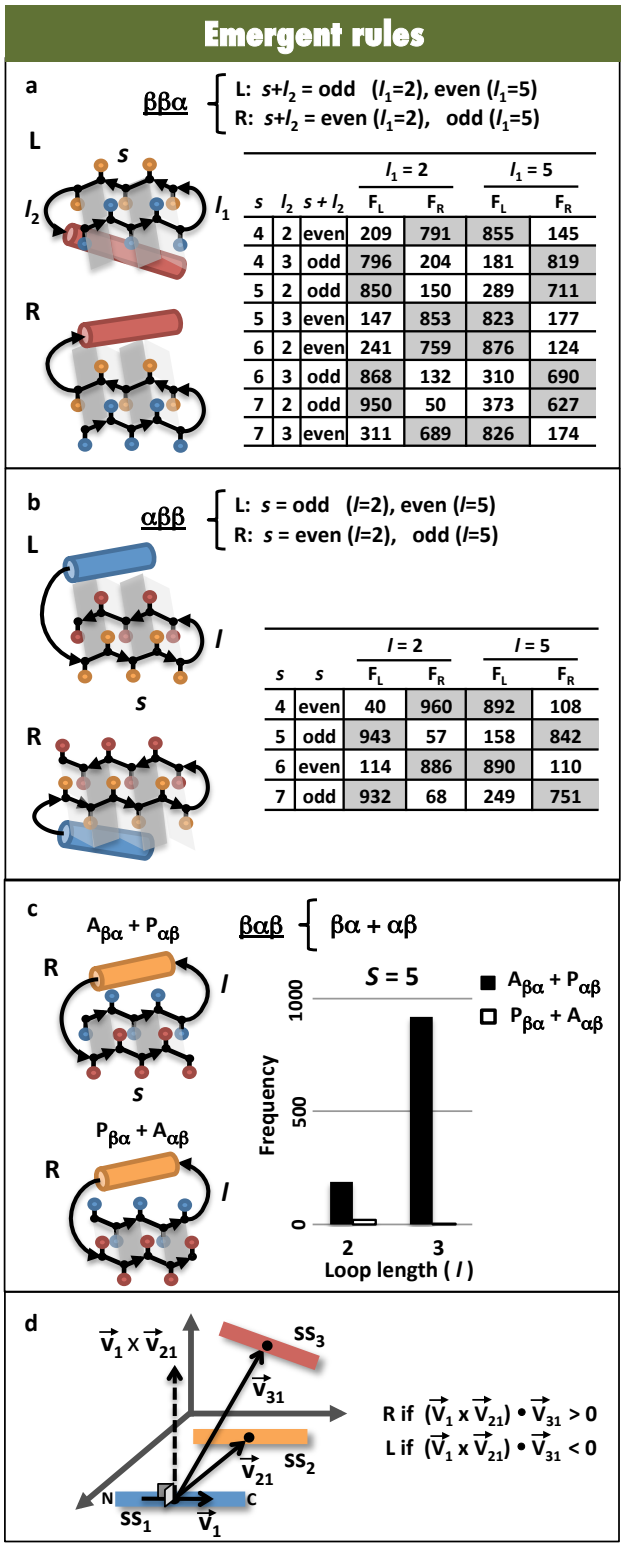
The rules are expressed as relations between the chirality of the unit and the secondary structure lengths. The chirality of the unit consisting of three secondary structure elements is defined as right handed (R) if the cross product of the vector along the axis of the first element with the vector from the center of the first element to the center of the second element points toward the third element, and left handed (L) otherwise (Supplementary Fig. 1d).

For ββα-units, both R- and L- topologies are observed in nature. Thus, negative design to destabilize one of the topologies is important. Secondary structure constrained folding simulations (see Supplementary Method 2) showed that the choice between R- or L- topologies depends on the strand and loop lengths, in particular, on l_1 and the sum $s+l_2$ (Supplementary Fig. 1a). A similar trend is also observed in native protein structures (Supplementary Fig. 5). This rule emerges from the combination of the ββ- and βα-rules. The ββ-rule relates the loop length l_1 to the pleating of the β-hairpin, and therefore the pleat direction at the end of the second strand depends on the strand length (assuming there are no beta bulges). Given the pleat direction, the helix direction is then determined by the loop length l_2 by the βα-rule (see Supplementary Fig. 6 for details).

For αββ-units, both R- and L- topologies are observed in nature. Thus, negative design to destabilize one of the topologies is again important. Secondary structure constrained folding simulations revealed that the choice depends on whether the strand lengths are even or odd and on the loop length (Supplementary Fig. 1b), and this trend is again observed in native protein structures (Supplementary Fig. 7). This rule emerges from the combination of the αβ- and ββ- rules. As in the case of ββα-units, the pleating of the β-hairpin is determined by the loop length by the ββ-rule. The pleat of the first residue in the strand immediately following the helix then depends on whether the strand has an even or odd number of residues, and this in turn determines the helix direction as specified by the αβ-rule (see Supplementary Fig. 8 for details).

βαβ-units are almost always right handed in protein structures[S1] and hence it is only necessary to focus on positive design for this motif. We performed secondary structure constrained folding simulations varying the strand, helix and loop lengths and measured the frequency of formation of a βαβ-motif in which the two strands make a parallel strand pairing and the helix packs on the both strands. We found that the frequency of formation of the motif depends on the loop lengths according to the βα- and αβ- rules (Supplementary Fig. 1c and 9), and also that the strand and helix lengths are codependent (Supplementary Fig. 10).

Supplementary Figure 1 | Emergent rules.

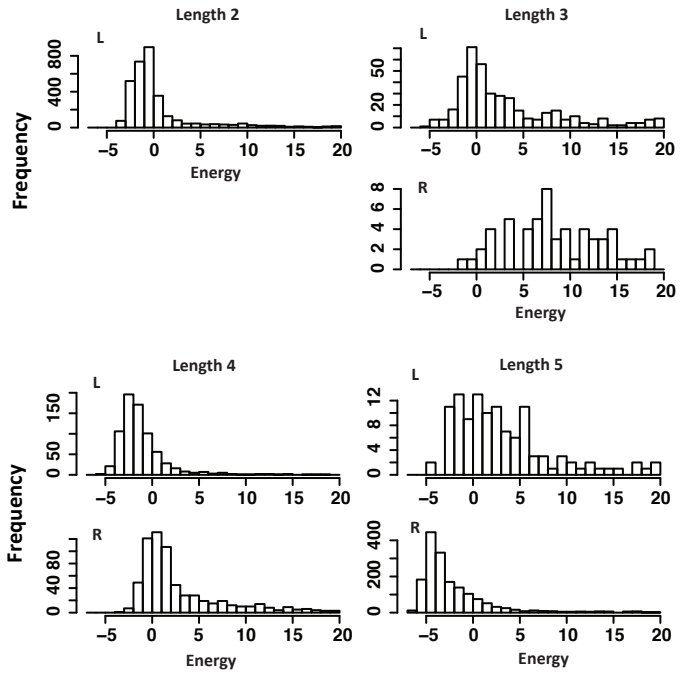


			$l_1 = 2$		$l_1 = 5$	
s	l_2	$s + l_2$	F_L	F_R	F_L	F_R
4	2	even	209	791	855	145
4	3	odd	796	204	181	819
5	2	odd	850	150	289	711
5	3	even	147	853	823	177
6	2	even	241	759	876	124
6	3	odd	868	132	310	690
7	2	odd	950	50	373	627
7	3	even	311	689	826	174

		$l = 2$		$l = 5$	
s	s	F_L	F_R	F_L	F_R
4	even	40	960	892	108
5	odd	943	57	158	842
6	even	114	886	890	110
7	odd	932	68	249	751

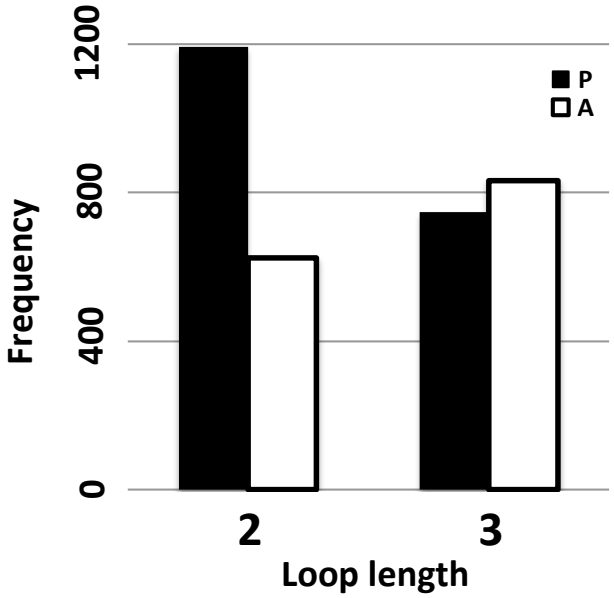
Supplementary Figure 1 | Emergent rules.

All data were obtained from Rosetta folding simulations. **a,** ββα-rule. L- and R- topologies of ββα-units are illustrated. The relationship between secondary structure lengths and the L- and R- topologies is given by the formula. The observed frequencies of L- and R- topologies in folding simulations for different secondary structure lengths are shown in the table; F_L (F_R) are the frequencies of the L- (R-) topology; shaded table elements satisfy the rule. Consistent with the rule, the illustrated L-topology has secondary structure lengths s=5, l_1=2, and l_2=2, while the illustrated R-topology has s=5, l_1=2, and l_2=3 (s: strand, l_1, l_2: loops). **b,** αββ-rule. As in **a**, the illustrated topologies have secondary structure lengths consistent with the formula: s=5 and l=2 for L, and s=6 and l=2 for R. The loop length immediately following the helix was fixed to 2 for all the simulations. **c,** βαβ-rule. Only the R-topology is consistent with the chirality of the polypeptide chain. The optimal secondary structure lengths for folding to the βαβ-topology follow from the βα- and αβ- rules. The two pleatings of the βαβ-topology for strand lengths 5 are illustrated: Aβα + Pαβ and Pβα + Aαβ. The observed frequency for each pleating in the folding simulations (l=2,3, helix length 14, and the loop length 2 following the helix) is shown in the histogram. The Pβα + Aαβ pleating that violates the αβ-rule was rarely observed, and the Aβα + Pαβ was much observed when l=3, which is consistent with the βα-rule. **d,** Chirality of three consecutive secondary structure elements. SS_1, SS_2 and SS_3 represent three secondary structure elements. $\overrightarrow{V_1}$ is a vector along the axis of the first element, $\overrightarrow{V_{21}}$ is a vector from the center of the first element to the center of the second element, and $\overrightarrow{V_{31}}$ is a vector from the center of the first element to the center of the third element. The chirality of the unit consisting of the three secondary structure elements is right handed (R), if $(\overrightarrow{V_1} \times \overrightarrow{V_{21}}) \cdot \overrightarrow{V_{31}} > 0$, and left handed (L), if $(\overrightarrow{V_1} \times \overrightarrow{V_{21}}) \cdot \overrightarrow{V_{31}} < 0$.



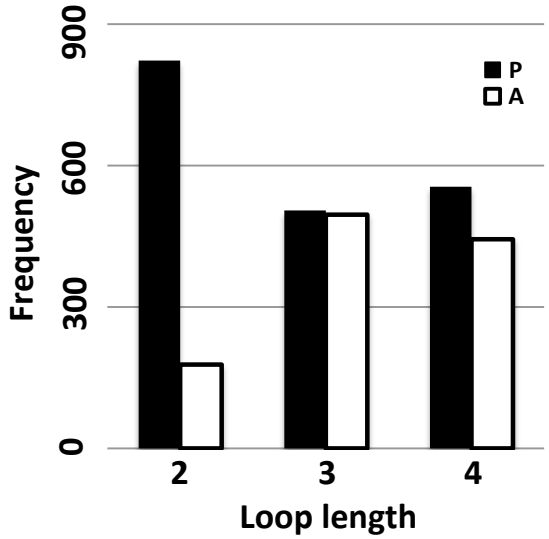
Supplementary Figure 2 | Torsion energies of loops are correlated with the chirality of β-hairpins in natural proteins.

The torsion energies for L- and R- hairpins were calculated for each loop length using Rosetta with the rama score term[S2] 1.0. There are no R-hairpins at the loop length 2. For 2 and 3 residue loops, L-hairpins have lower energy, and for 5 residue loop, R-hairpins have lower energy.



Supplementary Figure 3 | The βα-rule derives in part from the bendability of the backbone.

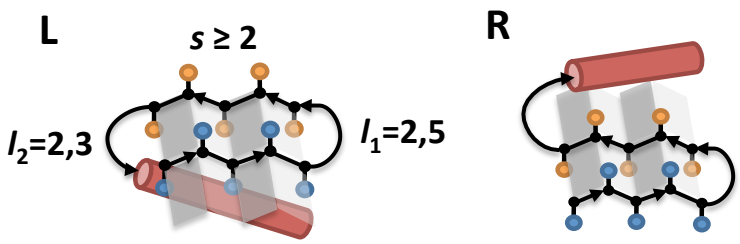
The right histogram of Fig. 1b shows the native structure distribution for the orientation of βα-units for which the angle between the helix and strand is ≤ 60°. The above histogram shows the native structure distribution without angle filtering. Since the loop length dependence is much stronger in Fig. 1b than in the histogram above, we infer that the rule arises in part from restriction imposed by chain bendability.



Supplementary Figure 4 | The αβ-rule for loop lengths 3 and 4 derives in part from hydrogen-bonded helix capping.

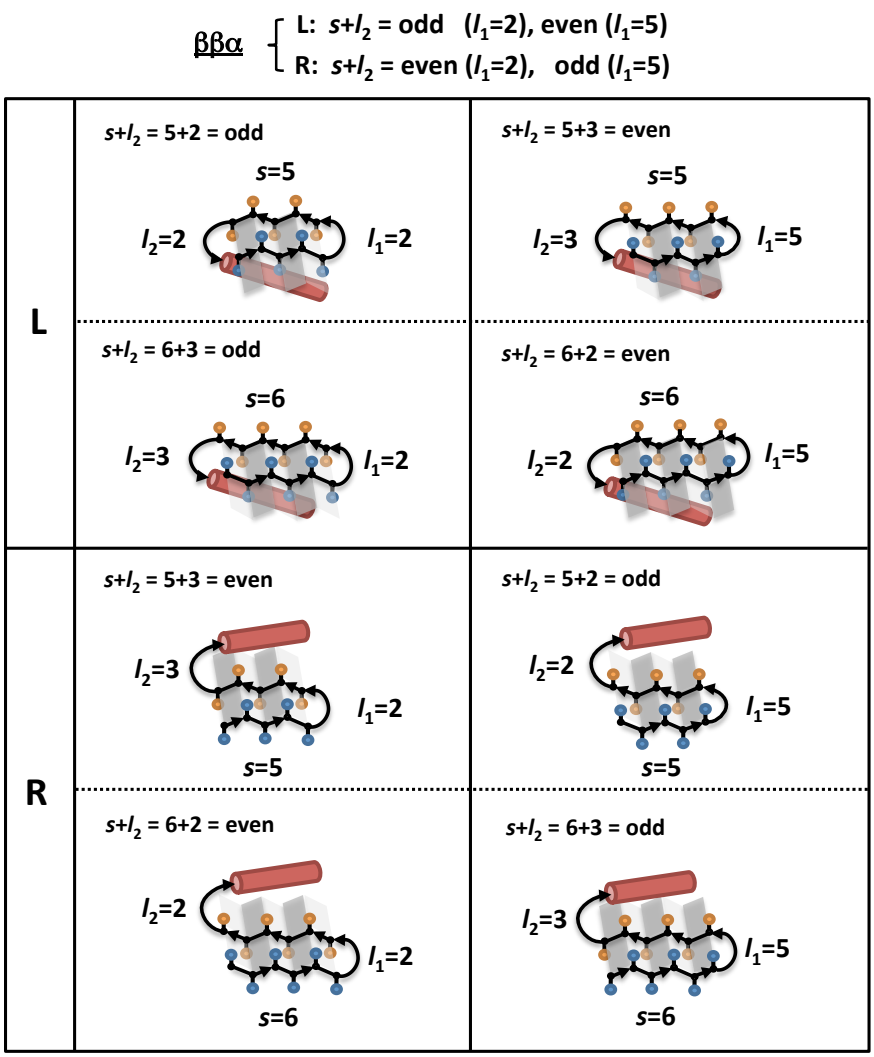
The left histogram of Fig. 1c shows the simulation structure distribution for the orientation of αβ-units for which the loop provides a hydrogen-bonded helix capping and does not extend the β-strand (see Supplementary Method 1). The above histogram shows the simulation structure distribution for all αβ-units. Since the trend that the preferred orientation of αβ-units is P is much stronger in Fig. 1c than in the histogram above for loop lengths 3 and 4, we infer that the rule for the loop lengths 3 and 4 arises in part from a hydrogen-bonded helix capping.

ββα	l_1 = 2		l_1 = 5	
$s + l_2$	F_L	F_R	F_L	F_R
even	2	20	22	8
odd	18	1	5 *	5 *



Supplementary Figure 5 | Dependence of ββα-unit chirality on secondary structure lengths in native structures.

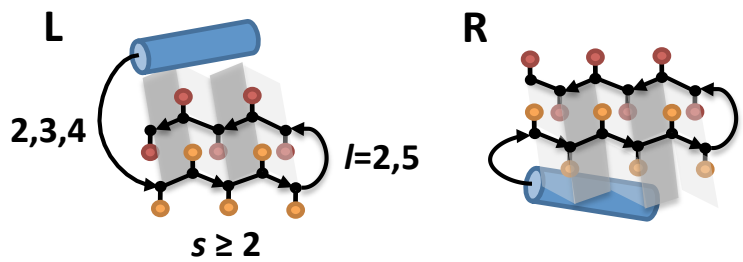
s represents the strand lengths, l_1 is the loop length between the strands, and l_2 is the loop length between the second strand and the helix. The table shows the frequencies of L- and R- topologies (F_L and F_R) in nature depending on the sum s+l_2 and l_1; shaded table elements satisfy the rule. We found that the native structures of ββα-units also follow the ββα-rule that we presented in Supplementary Fig. 1a for the simulations, except for l_1=5 and s+l_2 =odd (indicated by asterisks).



Supplementary Figure 6 | Explanation for ββα-unit emergent rule.

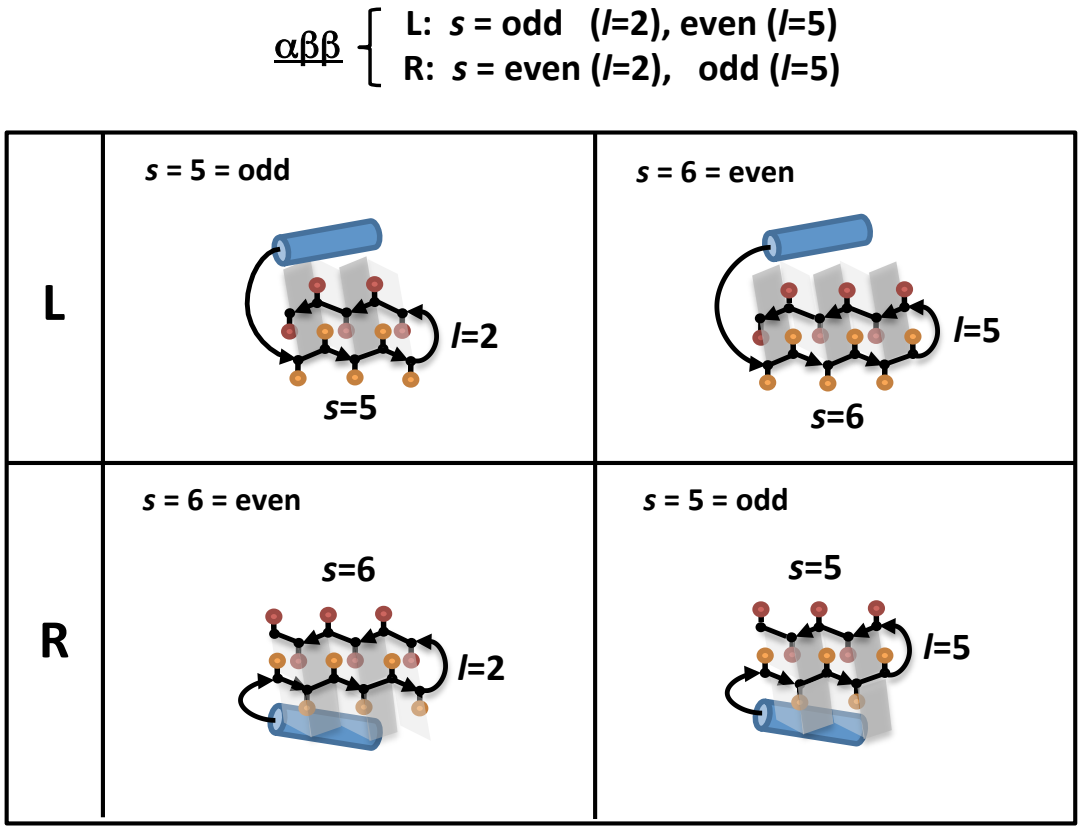
The dependence of the chirality of ββα-units on the secondary structure lengths is illustrated when the strand lengths s are 5 and 6. When the loop length l_1 is 2 (left boxes), the pleat of the strand residue preceding or following the loop points downward by the ββ-rule. The pleat direction at the end of the second strand is determined by the strand lengths: downward (s=5) and upward (s=6). The helix direction is then determined by the βα-rule: P if the loop length l_2 is 2, and A if l_2 is 3. Hence the chirality of the ββα-unit is L when $s+l_2$ = odd, and R when $s+l_2$ = even. When the loop length l_1 is 5 (right boxes), the pleating of the β-hairpin is flipped. Therefore, the $s+l_2$ dependence of the chirality of the ββα-unit is also flipped.

αββ	*l* = 2		*l* = 5	
s	F_L	F_R	F_L	F_R
even	19	30	21	4
odd	36	3	9	18



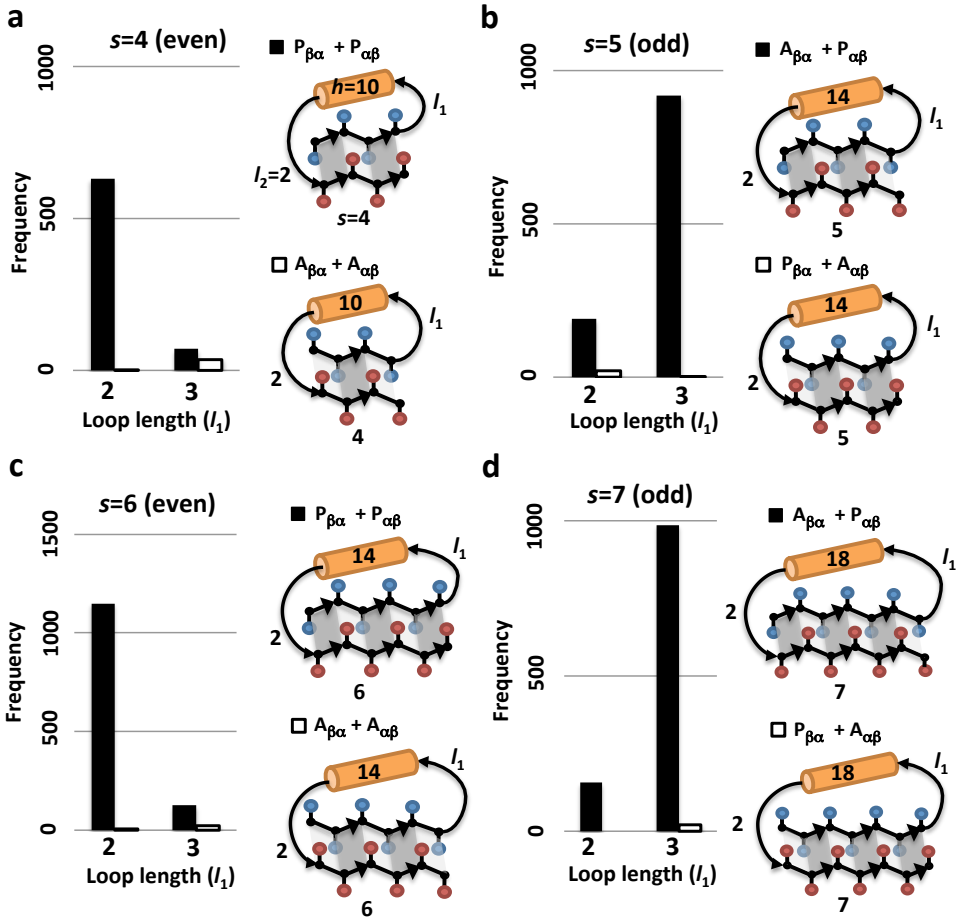
Supplementary Figure 7 | Dependence of αββ-unit chirality on secondary structure lengths in native structures.

s represents the strand lengths and *l* is the loop length between the strands. For the loop between the helix and the first strand, we took the structures that have the lengths, 2, 3 and 4. The table shows the frequencies of the L- and R- topologies (F_L and F_R) in nature depending on *s* and *l*; shaded table elements satisfy the rule. We found that the native structures of αββ-units also follow the αββ-rule that we presented in Supplementary Fig. 1b for the simulations.



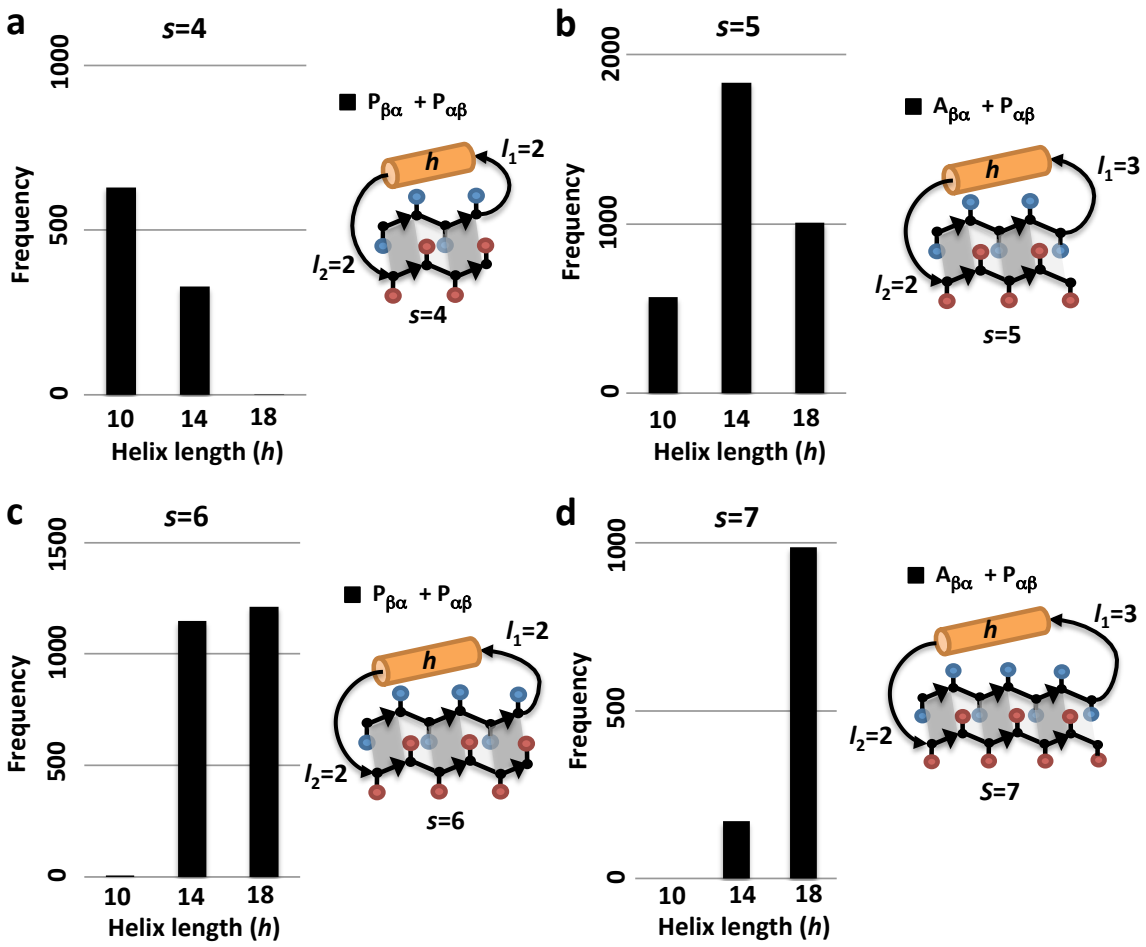
Supplementary Figure 8 | Explanation for αββ-unit emergent rule.

The dependence of the chirality of αββ-units on the secondary structure lengths is illustrated when the strand lengths *s* are 5 and 6. When the loop length *l* is 2 (left boxes), the pleat of the strand residue preceding or following the loop points downward by the ββ-rule. The pleat direction at the beginning of the first strand is determined by the strand lengths: downward (*s*=5) and upward (*s*=6). The helix direction is almost always P by the αβ-rule. Hence the chirality of the αββ-unit is L when *s* = odd, and R when *s* = even. When the loop length *l* is 5 (right boxes), the pleating of the β-hairpin is flipped. Therefore, the dependence of the chirality of the αββ-unit on the strand lengths *s* is also flipped.



Supplementary Figure 9 | Emergent rule for βαβ-unit (1).

Folding simulations of βαβ-units were carried out for different secondary structure lengths: loop lengths l_1=2,3 and l_2=2, and strand and helix lengths **a)** s=4, h=10, **b)** s=5, h=14, **c)** s=6, h=14, and **d)** s=7, h=18. For each strand length, there are two types of pleating of βαβ-motifs (illustrated on the right side) in which the two strands make a parallel strand pairing and the helix packs on the both strands. These pleatings are represented using the orientations (P and A) of βα- and αβ- units. Taking the structures that form a βαβ-motif at the end point, we counted the structures for each pleat type in the histograms. The simulation results show that βαβ-units follow directly from the βα- and αβ- rules. For all cases of the strand lengths, the structures with the sheet pleating including Pαβ are highly observed (filled bars), while those with the sheet pleating including Aαβ, which violate the αβ-rule, are rarely observed (open bars). In addition, when the strands have an even number of residues (**a, c**), Pβα + Pαβ is frequently observed when the loop l_1 is 2, which follows the βα-rule. When the strands have an odd number of residues (**b, d**), Aβα + Pαβ is frequently observed when the loop l_1 is 3, which follows the βα-rule.



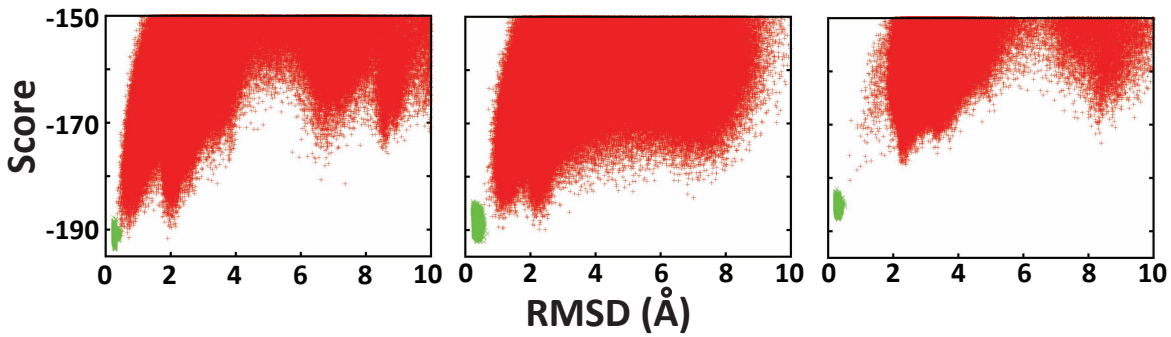
Supplementary Figure 10 | Emergent rule for βαβ-unit (2).

Folding simulations of βαβ-units were carried out for different helix lengths h=10, 14, 18. The other secondary structure lengths were fixed to be optimal (see Supplementary Fig. 9): **a)** l_1=2, l_2=2, s=4, **b)** l_1=3, l_2=2, s=5, **c)** l_1=2, l_2=2, s=6, and **d)** l_1=3, l_2=2, s=7. Taking the end point structures of the simulations, we counted the structures that form the βαβ-motif following the βα- and αβ- rules. The simulation results show the codependency between the helix and strand lengths: the longer strand favors the longer helix.

Supplementary Discussion 2 | Reasons for choosing the five specific folds.

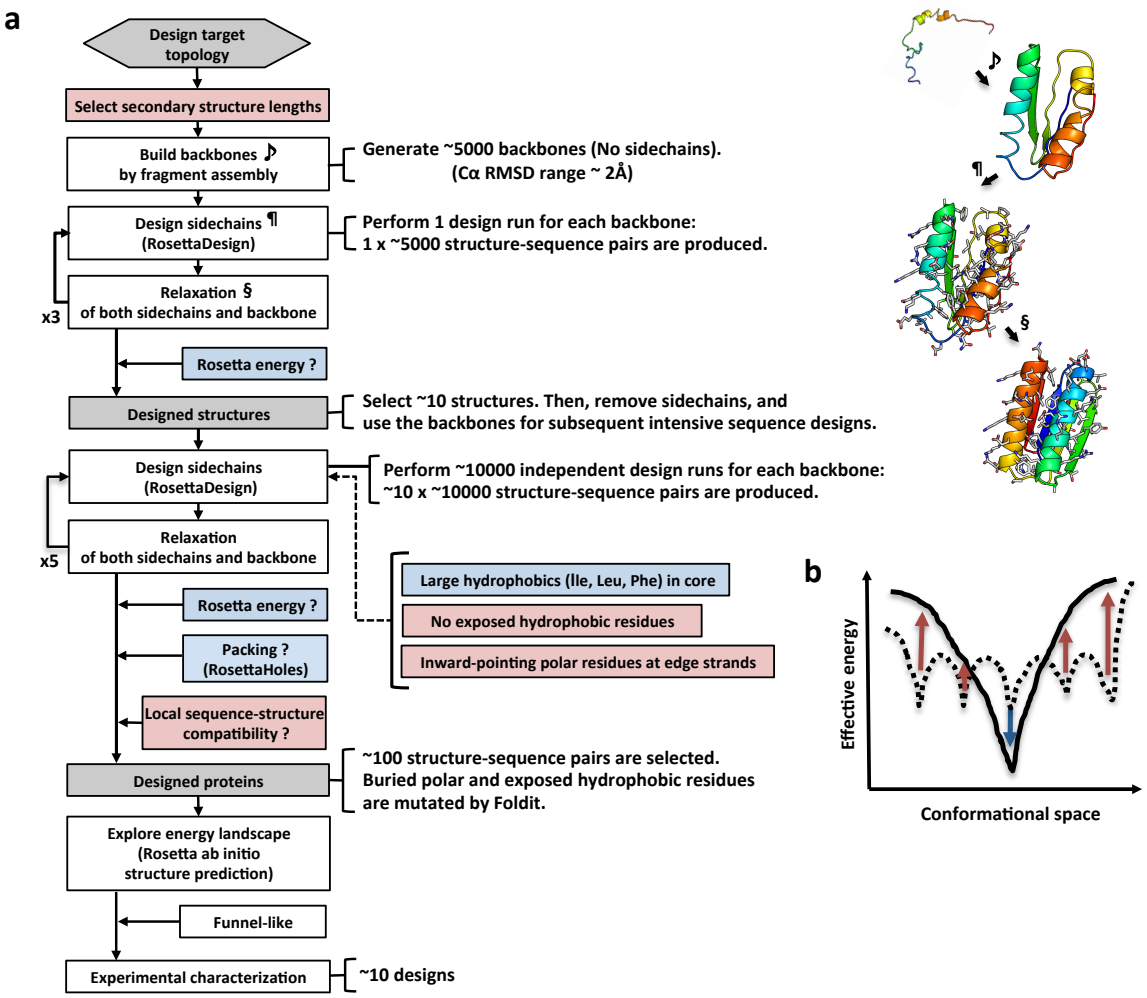
We chose a diverse set of folds that represent the variation in alpha-beta proteins found in nature; **Fold-I**: Ferredoxin-like fold. Two helices pack against an antiparallel β-sheet. **Fold-II**: Rossmann2x2 fold. Two helices pack against a parallel β-sheet from both sides of the sheet. **Fold-III**: IF3-like fold. Two helices pack against a β-sheet, same as Fold-I, but the β-sheet has the mixture of parallel and antiparallel strand-pairings. **Fold-IV**: Ploop2x2 fold. Two helices pack against a parallel β-sheet from both sides of the sheet, same as Fold-II, but the order of the two center strands is switched. **Fold-V**: Rossmann3x1 fold. Three and one helices pack against a parallel β-sheet from both sides of the sheet.

Folds-I and III are α+β proteins and helices are packed against a sheet from one side of the sheet. On the other hand, Folds-II, IV, and V are α/β proteins and helices are packed against a sheet from both sides of the sheet. Fold-I, and Fold-II and V are superfolds that are frequently observed in nature. Folds-III and IV are less frequently but still occasionally observed in nature. Fold-III is different from the other folds in that the sheet has the mixture of parallel and antiparallel strand-pairings. Fold-IV is a challenging target to design since it differs from Fold-II only in the internal strand swap.



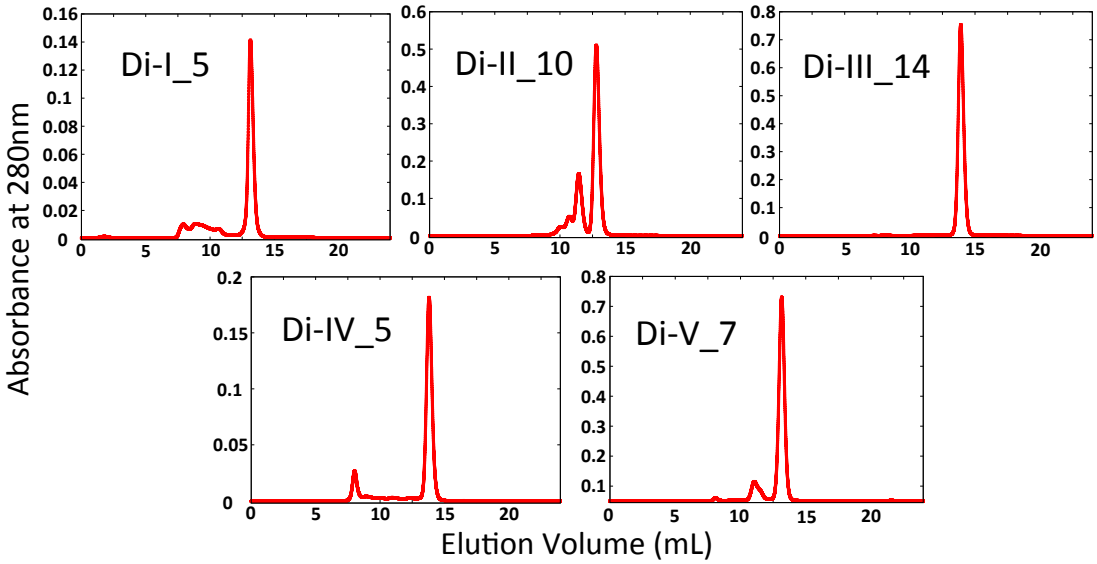
Supplementary Figure 11 | Examples of non-funneled energy landscapes.

Non-funneled energy landscapes of designed sequences for Fold-III obtained from Rosetta ab initio structure prediction simulations. Red points represent the lowest energy structures obtained in independent Monte Carlo structure prediction trajectories on Rosetta@home starting from an extended chain; the y-axis is the Rosetta all atom energy, the x-axis, the root mean square deviation (RMSD) to the design model. Green points represent the lowest energy structures obtained in trajectories starting from the design model. The energy distributions of green points are similar to or lower than that of the best design for Fold-III, Di-III_14 (Fig. 3 row(3) column(a)).



Supplementary Figure 12 | Computational protocol for designing ideal protein structures with funneled energy landscapes.

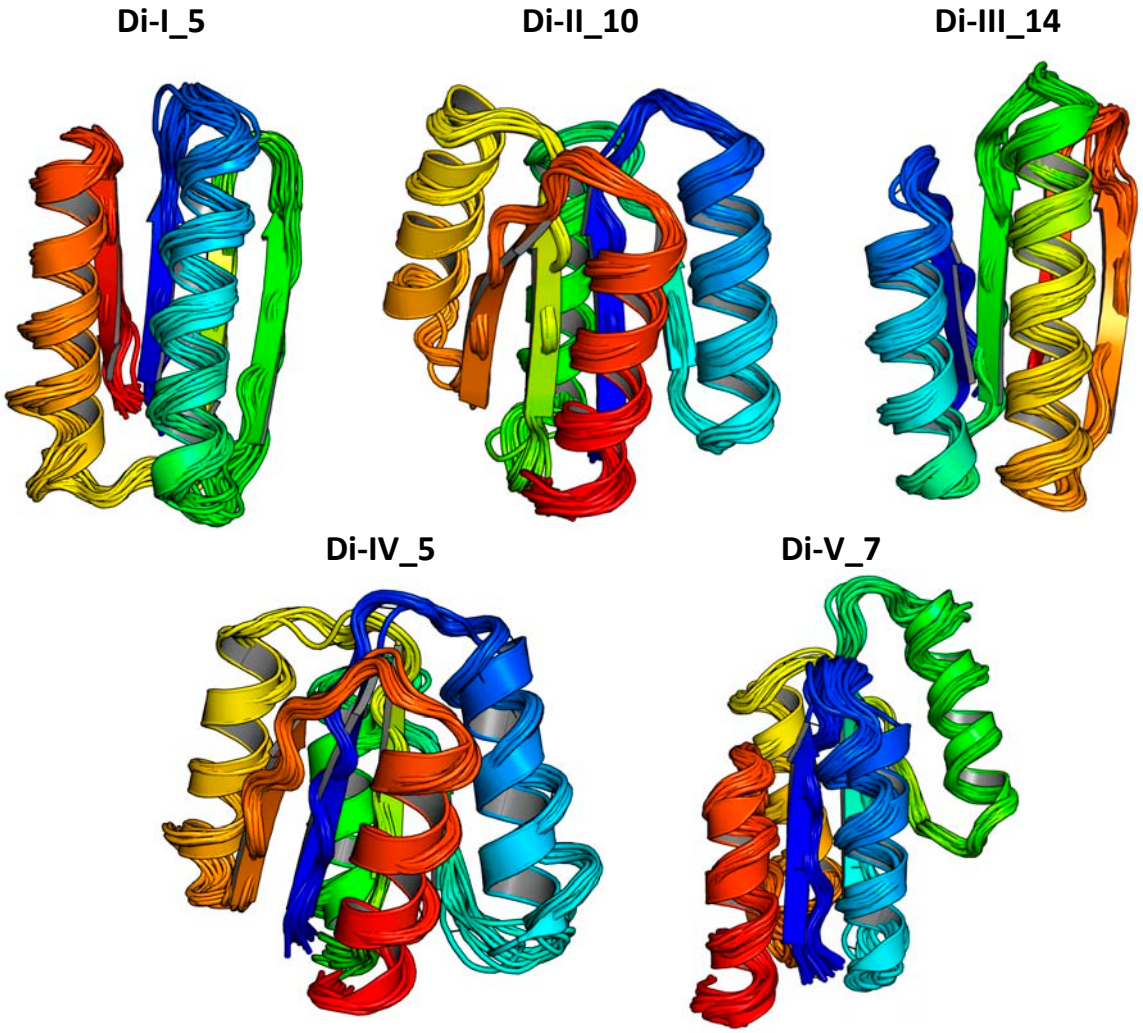
a, Flowchart of the design protocol. The red boxes indicate negative design steps, and the blue boxes represent positive design steps. An example of design process of Fold-I is illustrated in the right figures; each symbol in the figures corresponds to the design process with the same symbol in the flowchart. See the references: RosettaDesign[S3], Relaxation of both sidechains and backbone[S4], RosettaHoles[S5], Foldit[S6], and Rosetta abinitio structure prediction[S2]. **b**, Illustration of a non-funneled energy landscape (dotted line), and a funneled energy landscape (solid line). The blue arrow represents positive design, and the red arrows indicate negative design. ** The command lines and the examples of input files are available from the Rosetta SVN code repository in the directory: https://svn.rosettacommons.org/source/trunk/rosetta/rosettta_demos/public/ideal_proteins/.



	theoretical MW (kDa)	predicted MW (kDa) (% of main peak)
Di-I_5	10.2	12.9 (74)
Di-II_10	13.1	13.8 (67)
Di-III_14	9.8	9.5 (100)
Di-IV_5	12.7	12.9 (90)
Di-V_7	13.1	12.5 (88)

Supplementary Figure 13 | Oligomerization state of design for each of the five folds by SEC-MALS.

The volume 100 μl of 400-700 μM protein samples was injected into a Superdex 75 10/300 GL column equilibrated with PBS buffer (pH 7.4). The absorbance at 280 nm and the light scattering data at 658 nm were collected along the elution volume, which were analyzed to give the molecular weight (MW) of the main peak. The predicted MW by the analysis is presented in the third column in the table with the percentage of the main peak in parenthesis. The theoretical MW calculated from the design sequence is shown in the second column.



Supplementary Figure 14 | Superpositions of computational models and NMR structures.

Computational models are shown in cartoon and NMR structures, in wireframe.

	Expressed	Soluble	Alpha-beta protein CD spectrum	*Tm* (°C)	Monomeric	Well resolved 1D-NMR
1	✔	✔	✔	48	*	✔
2	✔	✔				
3	✔	✔	✔	74	§	✔
4	✔	✔				
5	✔	✔	✔	>> 95	✔	✔
6						
7						
8	✔	✔	✔	56	✔	
9	✔	✔	✔	> 95		
10	✔	✔	✔	> 95		
11	✔					

Supplementary Table 1 | Summary of experimental results of 11 designs for Fold-I.

Each row corresponds to the results for each design. The columns give the results for each criterion, of which the details are described in Table 1. The design that satisfies a criterion is shown with a check mark and the design that does not satisfy a criterion is shown in white blank. The case that the experiment was not conducted is shown in gray blank.

* The molecular weight of design #1 could not be obtained by SEC-MALS because the design has no TRP and TYR and the concentration could not be calculated from the absorbance at 280 nm.

§ The main peak of the absorbance at 280 nm was dimeric state.

	Expressed	Soluble	Alpha-beta protein CD spectrum	*Tm* (°C)	Monomeric	Well resolved 1D-NMR
1	✔	✔	✔	>> 95	*	
2	✔	✔	✔	>> 95		
3	✔	✔	✔	>> 95	§	
4	✔	✔	✔	>> 95		
5	✔	✔	✔	>> 95	✔	✔
6	✔	✔	✔	> 95		
7	✔	✔	✔	> 95	✔	✔
8	✔	✔				
9	✔	✔				
10	✔	✔	✔	>> 95	✔	✔
11	✔	✔	✔	> 95	*	
12	✔	✔	✔	>> 95	✔	✔

Supplementary Table 2 | Summary of experimental results of 12 designs for Fold-II.

The summary was given in the same way as the Supplementary Table 1.

* The main peak of the absorbance at 280 nm was dimeric or trimeric state.

§ The main peak of the absorbance at 280 nm was dimeric state.

	Expressed	Soluble	Alpha-beta protein CD spectrum	*Tm* (°C)	Monomeric	Well resolved HSQC
1	✔	✔	*	> 95	✔	✔
2	✔	✔	✔	≈ 95		
3						
4	✔	✔	✔	88	§	
5	✔	✔	✔	≈ 95		
6	✔					
7	✔					
8	✔	✔	✔	60		
9	✔	✔	✔	≈ 95		
10	✔	✔	✔	> 95	✔	✔
11	✔	✔	✔	≈ 95	✔	✔
12	✔	✔	✔	73	✔	✔
13	✔	✔			✔	
14	✔	✔	✔	>> 95	✔	✔

Supplementary Table 3 | Summary of experimental results of 14 designs for Fold-III.

The summary was given in the same way as the Supplementary Table 1.

* The CD spectrum was not characteristic of alpha-beta proteins, but indicated that this protein forms some secondary structures.

§ The main peak of the absorbance at 280 nm was dimeric state.

	Expressed	Soluble	Alpha-beta protein CD spectrum	*Tm* (°C)	Monomeric	Well resolved HSQC
1	✔	✔	✔	≈ 95	✔	✔
2	✔	✔	✔	60	✔	
3	✔	✔	✔	78	✔	✔
4						
5	✔	✔	✔	≈ 95	✔	✔

Supplementary Table 4 | Summary of experimental results of 5 designs for Fold-IV.

The summary was given in the same way as the Supplementary Table 1.

	Expressed	Soluble	Alpha-beta protein CD spectrum	*Tm* (°C)	Monomeric	Well resolved HSQC
1	✔	✔			*	
2	✔	✔			*	
3	✔	✔	✔	≈ 95		
4	✔	✔				
5						
6	✔	✔	✔	≈ 95		
7	✔	✔	✔	> 95	✔	✔
8	✔					
9	✔	✔			*	
10	✔	✔			*	
11	✔	✔			*	
12	✔	✔			*	

Supplementary Table 5 | Summary of experimental results of 12 designs for Fold-V.

The summary was given in the same way as the Supplementary Table 1.

* The main peak of the absorbance at 280 nm was dimeric state.

	RMSD between design and NMR (Å)		ΔG (kcal/mol)	*m*-value (kcal/(mol•M))	*Tm* (°C)
	Cα atoms	Heavy atoms			
Di-I_5	1.2	2.2	9.1	2.0	>> 95
Di-II_10	1.1	1.9	14.9	3.2	>> 95
Di-III_14	1.1	2.1	5.6	2.0	>> 95
Di-IV_5	1.7	2.3	4.8	3.3	≈ 95
Di-V_7	2.0	2.8	7.3	2.9	> 95

Supplementary Table 6 | Experimental data of design for each of the five folds.

The second and third columns show the averaged RMSD between the design model and the 20 NMR structures using Cα atoms and heavy atoms respectively. The computationally designed region corresponds to the region from N=2 Glu to N=77 Arg in the NMR structures (PDB code: 2kl8) for Di-I_5, N=2 Leu to N=100 Gly in 2lv8 for Di-II_10, N=3 Leu to N=74 Gly in 2ln3 for Di-III_14, N=2 Gly to N=102 Ala in 2lvb for Di-IV_5, and N=3 Ser to N=101 Gly in 2lta for Di-V_7. These regions were used for RMSD calculations. The subsequent columns show the free energy of unfolding ΔG, its dependency on the denaturant, *m*-value, and the melting temperature *Tm*.

	Di-I_5	Di-II_10	Di-III_14	Di-IV_5	Di-V_7
NMR distance and dihedral restraints					
Distance restraints					
Total NOE	1359	4081	2711	2835	2961
Intra-residue	371	782	576	515	754
Inter-residue					
Sequential (|i-j| = 1)	354	914	613	663	699
Medium-range (|i-j| ≤ 4)	227	959	577	642	577
Long-range (|i-j| ≥ 5)	407	1426	945	1015	931
Intermolecular					
Hydrogen bonds	64	88	70	82	38
Total dihedral angle restraints					
phi	64	78	54	80	82
psi	64	78	54	80	82
Total RDCs	0	126	104	0	99
Q(%, alignment media 1[§])		17.2	11.2		22.9
Q(%, alignment media 2[§])		13.9	9.5		16.4
Structure statistics					
Violations					
Distance restraints[¶] (Å)					
mean	0.000	0.001	0.001	0.001	0.001
rmsd	0.004	0.008	0.009	0.010	0.011
sd	0.004	0.008	0.009	0.010	0.011
Dihedral angle restraints (°)					
mean	0.078	0.130	0.148	0.099	0.122
rmsd	0.310	0.618	0.632	0.490	0.625
sd	0.300	0.604	0.614	0.480	0.613
Max. distance restraint violation (Å)	0.17	0.46	0.51	0.47	0.46
Max. dihedral angle restraint violation (°)	3.1	7.1	5.7	5.7	9.2
Deviations from idealized geometry					
Bond lengths (Å)	0.010	0.019	0.018	0.018	0.019
Bond angles (°)	0.5	1.2	1.2	1.1	1.2
Impropers (°)	1.2	1.7	1.8	1.7	2.0
Average pairwise r.m.s.d.** (Å)					
Heavy	1.79±0.16	1.28±0.09	1.18±0.14	1.23±0.10	1.62±0.12
Backbone	0.93±0.13	0.60±0.08	0.59±0.12	0.67±0.08	0.87±0.12
RPF Scores					
Recall	0.96	0.96	0.97	0.97	0.98
Precision	0.90	0.98	0.97	0.96	0.97
F-measure	0.93	0.97	0.97	0.96	0.97
DP-scores	0.69	0.89	0.90	0.87	0.87
Structure Quality Factors - overall statistics scores (raw/Z-scores[⌘])					
Procheck G-factor (phi / psi only)**	0.00/0.31	0.19/1.06	0.00/0.31	0.05/0.51	0.01/0.35
Procheck G-factor (all dihedral angles)**	0.00/-0.00	0.15/0.89	0.07/0.41	0.02/0.12	-0.14/-0.83
Verify3D	0.47/0.16	0.55/1.44	0.40/-0.96	0.46/0.00	0.52/0.96
ProsaII (-ve)	0.95/1.24	1.43/3.23	0.86/0.87	1.07/1.74	1.18/2.19
MolProbity clashscore	15.31/-1.16	17.7/-1.51	13.6/-0.81	16.45/-1.30	21.0/-2.08
Ramachandran Plot Summary from Richardson Lab's Molprobity**					
Most favored regions (%)	97.2	98.8	99.5	98.2	98.2
Allowed regions (%)	2.7	1.2	0.5	1.8	1.8
Disallowed regions (%)	0.1	0.0	0.0	0.0	0.0

Supplementary Table 7 | NMR and refinement statistics for protein structures*.

* Analyzed for the 20 lowest energy refined structures of design for each of the five folds, which are deposited in PDB: Di-I_5 (PDB code: 2kl8), Di-II_10 (2lv8), Di-III_14 (2ln3), Di-IV_5 (2lvb), Di-V_7 (2lta), by using PDBSTAT and PSVS 1.4[S7,8].

§ PEG and phage were used as alignment media 1 and 2.

¶ Calculated by using sum over r^{-6}.

** Calculated among 20 refined structures for well-defined residues that have sum of phi and psi order parameters[S9] S(phi)+S(psi)>1.8[S7]. The well-defined residues of Di-I_5: 2-9, 12-28, 32-47, 49-79; Di-II_10: 1-48, 51-99; Di-III_14: 5-64, 67-73; Di-IV_5: 3-26, 29-49, 52-102; Di-V_7: 3-10, 13-99.

⌘ With respect to mean and standard deviation for a set of 252 X-ray structures with sequence lengths < 500, resolution ≤ 1.80 Å, R-factor ≤ 0.25 and R-free ≤ 0.28; a positive value indicates a 'better' score.

Supplementary Method 1 | Simulations for two secondary structure elements.

In the simulations of ββ-units, the following secondary structure assignments were given as input: S_1[6]-L[2-5]-S_2[6] (L is loop, S is strand, and the numbers within the brackets are their lengths. L[2-5] means that the loop length ranges from 2 to 5). Multiple independent Monte Carlo trajectories were carried out, and then for the end point of each trajectory, we calculated the secondary structure using DSSP[S10]. The distribution of chirality in the end point structures, in which S_1 and S_2 made an antiparallel strand pairing, and the residue positions of the loop and the loop length exactly agreed with the input assignment, was then calculated (left in Fig. 1a).

βα- and αβ-units were studied in parts of βαβ-units in which the strands made a parallel strand pair with one another. Inputs of secondary structure assignments of the βαβ-units were S_1[6]-L_1[2-3]-H[14]-L_2[4]-S_2[6] for βα-units and S_1[6]-L_1[4]-H[14]-L_2[2-4]-S_2[6] for αβ-units. As for the ββ-units, multiple independent Monte Carlo trajectories were carried out for both cases, and the secondary structures of the end point structures were assigned by DSSP. We selected end point structures in which S_1 and S_2 made a parallel strand pairing and the residue positions of L_1 (L_2) and the L_1 (L_2) length were in agreement with the input assignment for βα-units (αβ-units). Finally, we computed the orientation of βα-units using the segments of S_1, L_1, and H of the βαβ-units (left in Fig. 1b), and computed the orientation of αβ-units using H, L_2, and S_2 of the βαβ-units (Supplementary Fig. 4). For the αβ-units, we further selected structures in which the loop provided the helix capping and did not extend the strand. These features were defined using the loop torsion angles described by the ABEGO letters[S11]; each letter represents a torsion space of phi, psi, and omega: A represents $-180^o \leq \varphi < 0^o$ and $-75^o \leq \psi < 50^o$; B is $-180^o \leq \varphi < 0^o$ and ($50^o \leq \psi < 180^o$ or $-180^o \leq \psi < -75^o$); E is $0^o \leq \varphi < 180^o$ and ($100^o \leq \psi < 180^o$ or $-180^o \leq \psi < -100^o$); G is $0^o \leq \varphi < 180^o$ and $-100^o \leq \psi < 100^o$; A, B, E, and G have $\omega \sim 180^o$; O has $\omega \sim -180^o$. We selected structures in which the loop provided the helix capping: GZ for 2-residue loop, GZX and ZGZ for 3-residue loop, and GZXX, ZGZX, AZGZ and BZAZ for 4-residue loop (Z = A, B; X can be any letter of the five letters)[S12]. We further eliminated structures in which the loop extended the strand: XBB for 3-

residue loop, and XXBB for 4-residue loop. Finally, we calculated the orientation of αβ-units using these filtered structures (left in Fig. 1c).

Supplementary Method 2 | Simulations for three secondary structure elements.

To study ββα-units, we considered ββαβ secondary structure string: S_1[4-7]-L_1[2,5]-S_2[4-7]-L_2[2,3]-H[14]-L_3[4]-S_3[6] (S_1 and S_2 have the same length). To study αββ-units, we considered βαββ secondary structure string: S_1[6]-L_1[4]-H[14]-L_2[2]-S_2[4-7]-L_3[2,5]-S_3[4-7] (S_2 and S_3 have the same length). Multiple independent Monte Carlo trajectories were carried out for both cases, and secondary structures of the end point structures were assigned by DSSP. For ββαβ units, we selected end point structures in which 1) S_1 and S_2 made an antiparallel pairing, 2) S_3 made an antiparallel pairing with S_1 for L topology or a parallel pairing with S_2 for R topology, and 3) the residue positions and lengths of L_1 and L_2 were in agreement with the input assignment. For βαββ units, we selected structures in which 1) S_2 and S_3 made an antiparallel pairing, 2) S_1 made an antiparallel pairing with S_3 for L topology or a parallel pairing with S_2 for R topology, and 3) the residue positions and lengths of L_2 and L_3 were in agreement with the input assignment. Finally, we calculated the chirality of ββα-units using S1, S2, and H, and the chirality of αββ-units using H, S2, and S3.

βαβ-units were studied with input secondary structure assignments of S_1[4-7]-L_1[2,3]-H[10, 14, 18]-L_2[2]-S_2[4-7] (S_1 and S_2 have the same length). We performed 8000 independent trajectories, and assigned the secondary structures of the end point structures by DSSP. We selected structures in which S_1 and S_2 made a parallel strand pairing, and all the secondary structure elements, S_1, L_1, H, L_2, and S_2, were in agreement with the input assignment.

Supplementary Method 3 | Analysis of natural protein features.

The PISCES server[S13] was used to collect 6875 X-ray structures in the PDB with resolution ≤ 2.5Å, R-factor ≤ 0.3, sequence lengths from 40 to 10000, and ≤ 25% sequence identity, and their secondary structures were assigned using DSSP. ββ-, βα-, αβ-, ββα-, and αββ- units were identified as adjacent secondary structure elements (no intervening secondary structure element other than a loop) in which the strands were at least 2 residues and the helices were at least 5 residues in length. For βα-, αβ-, ββα-, and αββ- units, the angle between a vector along the helix and a vector along the strand must be ≤ 60°. For βα- and ββα- units, we define the vector from the N (backbone amide nitrogen) atom of the first helix residue to the C (backbone carbonyl carbon) atom of the fourth helix residue as the helix vector, and the vector from the N to C atoms of the last residue in the strand immediately preceding the helix as the strand vector. For αβ- and αββ- units, we define the vector from the N atom of the fourth helix residue from the last to the C atom of the last helix residue as the helix vector, and the vector from the N to C atoms of the first residue in the strand immediately following the helix as the strand vector. The ββα- and αββ- units for which the strands include bulges were omitted.

Supplementary Method 4 | Definition of the chirality of a ββ-unit in fundamental rules.

In the Methods Summary, the chirality of a ββ-unit was considered using the vector $\vec{u}$ along the axis of the first strand, and the vector $\vec{v}$ perpendicular to $\vec{u}$ between the centers of the two strands. Since twisted strands lead to inaccurate assignments of the chirality, however, we used atom coordinates close to the loop between the strands for the definition of $\vec{u}$ and $\vec{v}$: $\vec{u}$ is a vector from the N (backbone amide nitrogen) to C (backbone carbonyl carbon) atoms of the strand residue preceding the connecting loop and $\vec{v}$ is a vector from the Cα atom of the strand residue preceding the loop to the Cα atom of the strand residue following the loop.

Supplementary Method 5 | Definition of the orientation of βα- and αβ- units in fundamental rules.

In the Methods Summary, the orientation of βα- and αβ- units was defined using the vector between the centers of the strand and helix. Since twisted strands and kinked helices lead to the inaccurate assignments of the orientation, however, we used a vector close to the loop for the definition of the vector. For a βα-unit, we used a vector from the Cα atom of the last residue in the strand to the average coordinate of the first 11 backbone atoms (N, C, and Cα) in the helix. For a αβ-unit, we used a vector from the average coordinate of the last 11 backbone atoms (N, C and Cα) in the helix to the Cα atom of the first strand residue. We defined the orientation of βα- and αβ- units as parallel (P) when the angle between the vector and the $\overrightarrow{C\alpha C\beta}$ vector of the strand residue closest to the helix is ≤ 80°, and antiparallel (A) when the angle is ≥ 100°.

Supplementary Table 8 | Designed sequences

Computationally designed sequences are shown in uppercase and residues added to allow expression, purification, and the spacer between the designed sequence and the N-terminal Met or the C-terminal His-tag are shown in lowercase.

Di-I_1	mEMDVRFRMNDLDGLIKAADEMKREAERANGTITKTLDGNDLEIRITNVTEQERKEIAQRAEELAKEFNGTVTKTIRgslehhhhhh
Di-I_2	mTLDIRLKMNDKDGLIDAAQRLEEEARKHKGTITKTIDGNEMEFRMTNMTEQFRKELMEEAEKLAKRYQGTITKTIRgslehhhhhh
Di-I_3	mEMDIRFRGDDLEAFQKAAERAEKEAKKYAGTVTKTLDGNDLEIRITGVTEEVRKELIQRAEELAKEFNITVTKTIRgslehhhhhh
Di-I_4	mELDIRFRGNDDEALKRAEEEMEEDAKKAAGTITKTVDGNDVEIRITNITQQWAERIRKEAERRAREEGTTATKTWRgslehhhhhh
Di-I_5	mEMDIRFRGDDLEAFEKALKEMIRQARKFAGTVTYTLDGNDLEIRITGVPEQVRKELAKEAERLAKEFNITVTYTIR(gs)lehhhhhh
Di-I_6	mKLNVRIRDSDQNKLQKALKKFIELARKSNGTITKTYTGTDLEIQITNITQEEIQKIAEEAQKLAQKINGTVTKTFTgslehhhhhh
Di-I_7	mTMEMRFRGDNQEGFDKAKDLAKKWAQKFNGTVTKTLTGNDLDIRITNVPEEARKKFKQWAEELAKKFNITVTKTITgslehhhhhh
Di-I_8	mTLDIKFHGDSPEAWEKAREMAEELAREFNGTVTKTITGDDVEIRITNIPEEAKQRGRERAEELAKEANITVTKTITgslehhhhhh
Di-I_9	mEAEVRFRMDDYDGLVKAIQRMIEEAKRANGTITKTITGDDVDIRITNITEREAKEIFREAQRLAQEFNGTVTKTYTgslehhhhhh
Di-I_10	mEMDIRFDGDDLEAFQEFLQRAKERAEKYAGTMTKTLDGNDLEIRITGVPEQARKEFAREAEELAREKNITVTKTIRgslehhhhhh
Di-I_11	mELDIRLKMNDEEGLKKARERMEEEARKHNGTFTKTLRGNDLELRMKNMTEKFRKELVEEAEKLAKRFNGTITKTFRgslehhhhhh

11 designs for Fold-I. The C-terminal gs spacer of Di-I_5 (shown in parenthesis) was removed in the NMR structure determination.

Di-II_1	mNYFILVFTNNEDIIREVERMAQDSGLQYRTVKSKDEAKKYLEEFRRRSQNIYVLLVVSDEKYLRELEELARKFDIQVTS VKAESPDKARDDVKEYSEKGgslehhhhhh
Di-II_2	mNYFILVFTNNKRIIEEVEEIAKKKGFQVRQVKDKDEARKYLEEFKKKSKNIYVIIIVGTEKYLRELKRFADEFDIQVRTRK VTSPDQARDDVEELSERFgslehhhhhh
Di-II_3	mNIFVLVFTSDEELIKYVEEMAKKDNVTVKHVKTKNEAKKYLDEFKKKSQNIFVILLVRDEKSWREFEELARKLNIQVRT SKIESPDKAKDSLRKFYEEFgslehhhhhh
Di-II_4	mNIFVLVFTTDKRLIEKIREIVEKQNTQVRTVESEDHAKDYLEEFRRRSKNIFVVLMVHTEEYLRRLKELAEKFEIDVTSM RVTSPDEAKDSVKDLIDKGgslehhhhhh
Di-II_5	mNIFVLVFTDDEKAYREIEKEVRRRGAEIRRVKDSEEARRYLEEFLRKDKNVYVIILVKNEEELEKFRELADKFNIQVRSR RVHSPEEAKRWFKELEKRFgslehhhhhh
Di-II_6	mIFVLVFWGTDEDAREEAEKVARKQNVKYRTVTTKDTMKDLFEKFKKESENIYVVIIVSTEEDLKKAKELAEEMDIQVR TRSARSPEEARKWAKKLLEEGgslehhhhhh
Di-II_7	mLIYVIIFGGSDELRDKVKELAKRKGAQVRTVHSKDELKKLLEEFKKQGDNVFVLILVNDEKMRKLAEELAKKYNLQIRT RRVQSPDIAKDRVKEYIEKVgslehhhhhh
Di-II_8	mLSYIFVFTDDRKLYEEAEKMARKQGFQLTRVETEDHFEKKLRELKKRSKNIYVLIIVKDRESLDKFKERAEESDVQVKS VEAQSPDEAKDWVKEYSEEIgslehhhhhh
Di-II_9	mIIYVLIITTNKKLIEEAEKMAKKANLELRTVKDDDEFKKYLDEFRDKDENIYVLVIVSTQEKLQKARKRAKEEEIDIRTRD AQSPDKAKDYIEKYFRKIgslehhhhhh
Di-II_10	mLLYVLIISNDKKLIEEARKMAEKANLELRTVKTEDELKKYLEEFRKESQNIKVLILVSNDEELDKAKELAQKMEIDVRTR KVTSPDEAKRWIKEFSEEGgslehhhhhh
Di-II_11	mLEYVIILGGSKKLIEEVKKLAENKGMEIRTVTSKDELKDLLKEFKEKSQNLFVIIIVRDEELLKKAEELAKKYNLQIRTYRA QSPDKAKDYVKEYYERIgslehhhhhh
Di-II_12	mILYVLIFTNDEKLIRKAKEMAEKMGIELRTVKDTNELKRYLEEFKRKDDNIQVIIIVTNDEDLKKATKLAREYNIDVRTR RATSPDEAKDLIKKYFEKGgslehhhhhh

12 designs for Fold-II.

Di-III_1	MQLKFTSNDENKMLQWMKDAIKQGKKLEFRFTSTDDDRIKKFLQLAEDLAKESGVQIKIKTKGDTYEVELEGslehhhhhh
Di-III_2	MQYKFTSSDTERMKQEMKDAIKNGKSLRFEFRSTDDDQLKKFLEEAEKLAKKSGVQIEIRWKGNTFEVQLTGslehhhhhh
Di-III_3	MKIKFKSDDENKIEKWLEEALKKGIEIEFRIRLNNDDRLDSIKDQFKKKVEEQGVQYEIRWEGNELRLEMKGslehhhhhh
Di-III_4	MQLRFKSNDKNQILKWMKEAIKKGIELEFEIESNDQNQLDEIKDEFEREVREQGVKYQIEEKGNKLELKVKGslehhhhhh
Di-III_5	MRYRFRSEDKNQILKWMEEAIKKGKEMDFEIDSNNDDQLDEIRDKFKDEVQKKGVEYKIETQGNTLRLIVKGslehhhhhh
Di-III_6	mKKLKFKSDDDNQIIKWMREAIKKGIKMKFEIEQTDDNRLEEIKRKFKDEVQKQGVEYKIEEKGNKLELEVKGslehhhhhh
Di-III_7	MQLRYRTQNEDQIKDLVKKAAQKGIQMEMQMQDNDKKQLEEMLKKVSEIAQKEGVQYQYSWQGDSLSVSVQGslehhhhhh
Di-III_8	MRLKYQSDDDNKMLQLMKDAIKNGKELEFKFTDTNDDQIKDFLKKAEDLARKSGVQIKLKTEGNDYEVNLRGslehhhhhh
Di-III_9	mDQYKFTSKDKDELLDWMKKMIQQGKRLEMEFRDTDDNKLKQFWEDIEREAKKQGVQIEYEQQGNTIEIRIQGslehhhhhh
Di-III_10	mDRLKITSNDKDELLERVKEAIEQGIELEIEIDDTNDDKIKEILDEFEKLVKKSGVQIEIRWQGNRLELEIRGslehhhhhh
Di-III_11	mSRIRIQTRDDEELRELVKRAAEQGIKVIIQIQDNDEKKLREIEEDAEKIARERGVQIKSRWQGSSLEIEIEGslehhhhhh
Di-III_12	MQLKFKSDDKDKMLQWMKDAIKQGKELEIEIDTNDDNRLDEMKDLAEDLARKQGVQIEIREQGNTIEVRLKGslehhhhhh
Di-III_13	mgLTRTIEDQDTKDLLEWLKKAIDDGKRLKIRFQDTDDNQLKEFEQRIEDLAKEKGVQIKKRTQGDKLEFELEGslehhhhhh
Di-III_14	mgLTRTITSQNKEELLEIALKFISQGLDLEVEFDSTDDKEIEEFERDMEDLAKKTGVQIQKQWQGNKLRIRLKGslehhhhhh

14 designs for Fold-III.

Di-IV_1	mGRVLLIVSTNKNDINQLKDLVRKSGPGKEVRTVSNSNQIRNVIQTAKSNGRPLIVFINGATDDDIKEFERDMQQEGLQ YRVVRSTDPEELRTEVKKFDNSDgslehhhhhh
Di-IV_2	mGKVLLIVSNDSNDITEVEREARKQGPGKETRTVTNKDDIETVINHMKNNGKPLIVFSSGATDQDIKYFEKVAQQSGVS YEVRKSQDPEELRTEVRNFVQSLgslehhhhhh
Di-IV_3	mGKVLLIVTSNQNLLNQIKKEIESQGPGKYTRTVTNSDDIRDVIKSARKSGGPMVVFNSGATDNDIKKFQSIASNEGIEYR VRTDTDPEELLSEARRFVKQAgslehhhhhh
Di-IV_4	mGKVLLIITTDSNILQKLRQRMEKSSPGKQSRTVTTDSDIRQAISNARQNGRPMVIFIRGGNSDRIDDFESIAKKEGITYD VVRNTDPEELRERVEEFVKNEgslehhhhhh
Di-IV_5	mGKVLLVISTDTNIISSVQERAKHNYPGREIRTATSSQDIRDIIKSMKDNGKPLVVFVNGASQNDVNEFQNEAKKEGVSY DVLKSTDPEELTQRVREFLKTAgslehhhhhh

5 designs for Fold-IV.

Di-V_1	mgSAIIIIYSTDDNKLLKWVKEVKDQGIEVYLLLSDDDEDRLKKWLDKLRSQGIEVREVKDDDDLKQILDDIKKKRPQLEIREVQSEDRMKKALESVEKSGslehhhhhh
Di-V_2	mgSIIVVIISSDDQKLKDWVEKVRRKGIEVIMIYKDKDQNRLDQVIKDMQNQGVEVRKVEDDDDLKEILDRIKKKRPQLEIREVQSEDRMKKALDEAEKRGslehhhhhh
Di-V_3	mgSIILVIYSSDKQNLEDKAKKVRKQGIEVFILLSDTDEQKLKDWLQKLRNQGLEVREVRDKNDLEQILKDIKKKRPQLELRKVTSEDRLKEVLDEAKKRGslehhhhhh
Di-V_4	mgSIIVVIYSSDKEELREKAEKARKQGLEVIILLSDDDKDRLEKKLEDLRKQGLEVREVRDDDDLKQILDDIRKKRPQLQIRDVQSEDRFKKVIKEAQERGslehhhhhh
Di-V_5	mgSAIIIIISSDQQELQDKAKKVREEGVEVIILLKDQDKKKLEEWLKKLRNQGLEVREVRDDDDMKQILKDFRKERPQLQIRTVKSDDRLKKALDDVKKNGslehhhhhh
Di-V_6	mgSVILVVISSDDEELRERAEKIRDQGIEVIVLLKDKDDDRLKDKIDKIKSQGVEVRQINDDDDLKKWLEEIKKKRPQLEIRKITDEDEFKKALEEAEKRGslehhhhhh
Di-V_7	mgSKIIVIISSDDTTLEELARKIKDEGLEVYILLKDKDEKRLEEKIQKLKSQGFEVRKVKDDDDIDKWIDKIKKERPQLEVRKVTDEDQAKQILEDLKKKGslehhhhhh
Di-V_8	mgSAIIIVYSTDDEKLLKEVKKAKDTGLEVFLLLSDNDDNRLDQWLKDLRSQGIEVRKVNDKNDLEKIIKDIKKKRPQLEVRKVTDTNQFEQILKDLKKKGslehhhhhh
Di-V_9	mgSVIIVVYSSDQENLEEIAQRIKDTGLEVIILLSDDDEQKLKEWLQKLRNQGVEVREVKDQNDLDDILDDIEKKRPQLKIRKVTDKQEAEDILRKAEKEGslehhhhhh
Di-V_10	mgSKIIIIYSSDDKTLLELVEKIKKTGLEVYLLLSDNDEQRLEEWLKKLRNQGFEVRKVNDKNDLEKIIKEIKKRRPQLEVRKVTDKNEAEDILKKLKKEGslehhhhhh
Di-V_11	mgSKIIVVYSSDKDKLKEIAEKIKETGLEVYILLSDTDEKKLKKWLDEIKSQGVEVREIRDDDDLKEWLDRIKKKRPQLEIREVTDKNQAEDILKKLKKEGslehhhhhh
Di-V_12	mgSAIVIIYSTDDEKLLKLVKKVKDKGLEVFLLLSNDDEQKLKEWLQKLRSQGIEVREIRDKNDLEEWIKRIKKKRPQLEVRKVTDKDEAEQILKDLEKKGslehhhhhh

12 designs for Fold-V.

References

S1 Richardson, J. S. Handedness of crossover connections in beta sheets. *Proceedings of the National Academy of Sciences of the United States of America* **73**, 2619-2623 (1976).

S2 Rohl, C. A., Strauss, C. E., Misura, K. M. & Baker, D. Protein structure prediction using Rosetta. *Methods Enzymol* **383**, 66-93, doi:10.1016/S0076-6879(04)83004-0 (2004).

S3 Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364-1368, doi:10.1126/science.1089427 (2003).

S4 Tyka, M. D. *et al.* Alternate states of proteins revealed by detailed energy landscape mapping. *J Mol Biol* **405**, 607-618, doi:10.1016/j.jmb.2010.11.008 (2011).

S5 Sheffler, W. & Baker, D. RosettaHoles: rapid assessment of protein core packing for structure prediction, refinement, design, and validation. *Protein Sci* **18**, 229-239, doi:10.1002/pro.8 (2009).

S6 Cooper, S. *et al.* Predicting protein structures with a multiplayer online game. *Nature* **466**, 756-760, doi:10.1038/nature09304 (2010).

S7 Bhattacharya, A., Tejero, R. & Montelione, G. T. Evaluating protein structures determined by structural genomics consortia. *Proteins* **66**, 778-795, doi:10.1002/prot.21165 (2007).

S8 Huang, Y. J., Powers, R. & Montelione, G. T. Protein NMR recall, precision, and F-measure scores (RPF scores): Structure quality assessment measures based on information retrieval statistics. *Journal of the American Chemical Society* **127**, 1665-1674 (2005).

S9 Hyberts, S. G., Goldberg, M. S., Havel, T. F. & Wagner, G. The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein science : a publication of the Protein Society* **1**, 736-751, doi:10.1002/pro.5560010606 (1992).

S10 Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637, doi:10.1002/bip.360221211 (1983).

S11 Kim, D. E., Blum, B., Bradley, P. & Baker, D. Sampling bottlenecks in de novo protein structure prediction. *J Mol Biol* **393**, 249-260, doi:10.1016/j.jmb.2009.07.063 (2009).

S12 Aurora, R. & Rose, G. D. Helix capping. *Protein Sci* **7**, 21-38, doi:10.1002/pro.5560070103 (1998).

S13 Wang, G. & Dunbrack, R. L., Jr. PISCES: a protein sequence culling server. *Bioinformatics* **19**, 1589-1591 (2003).