An evolution-based approach to *de novo* protein design and case study on *Mycobacterium Tuberculosis*

Pralay Mitra, David Shultis, Jeffrey R. Brender, Jeff Czajka, David Marsh, Felicia Gray, Tomasz Cierpicki, Yang Zhang

Supporting Information

Text S1: Structure feature predictions and DNA/protein sequences

The sequence search simulation in the evolution-based protein design approach (EBM) uses structural profile, secondary structure, solvent accessibility, and backbone torsional angle prediction terms, in addition to the physics-based force field from FoldX [1]. A number of sophisticated methods have been developed for predicting these features which all rely on the position specific scoring matrix (PSSM) from PSI-BLAST search [2-4]. However, these methods cannot be used in our design procedure due to the high CPU cost of PSI-BLAST, as the prediction is required at each step of the Monte-Carlo simulations. In the following, we describe the method development for single-sequence based structure predictions without using PSI-BLAST.

Secondary structure prediction. The secondary structure (SS) is predicted by back propagation neural network (NN) training methods[5], which considers three states: Helix (H), Sheet (E) and Coil (C). The NN training features are based on three fingerprints of secondary structure propensity score, amino acid composition score, and BLUSOM62 substitution score[6].

Secondary structure propensity score. The SS propensity score of an amino acid type x in a particular secondary structure state h is calculated from the statistics of 45,397 non-redundant protein structures from the Protein Data Bank (PDB)[7], i.e.

$$s(x,h) = \frac{C_{x,h}/C_h}{C_x/C} \qquad \text{(S1)}$$

where C is the total number of amino acid residues in the dataset, C_x is that with amino acid type x, C_h is that in secondary structure state h, and $C_{x,h}$ is that with amino acid type x and secondary structure state h. The first three cells (Index 1-3 in Fig S2) of the fingerprints are filled with the propensity score.

Amino acid composition Score. The amino acid composition score describes the sequence environment feature of amino acids. For a given amino acid type x at position i on the training protein set, we count the frequency of occurrence of other amino acids y_i, $f(x_i,y_i)$, within a sliding window (say within ±3 residues, as shown in Figure S1). The composition score $c(x_i)$ is then calculated as the frequency multiplied by the solvent accessibility of that residue, $SA(y_i)$, in a tri-peptide (A-X-A) conformation [8], i.e.

$$c(x_i) = f(x_i, y_i)SA(y_i) \qquad \text{(S2)}$$

Thus, each position has 20 composition score corresponding to each amino acid that constitutes a fingerprint of size 20 (Index 5-24 in Fig S2).

Neural network training. The single-sequence based secondary structure prediction is trained by a one-layer neural network with fingerprint features including SS propensity score, amino acid composition score, and BLOSUM62 substitution matrix (Figure S2). The NN was trained on 5,527 non-homologous proteins, with true secondary structure assigned by DSSP [9]. Three predictors were trained using different window size of 16, 17 and 21, where in the latter two the amino acid composition score was turned off. The final SS prediction is obtained by summing up the probability score of the three predictors, where the state of the highest probability score is returned. The test on 625 proteins non-redundant to the training set shows that the overall Q3 accuracy is 69.3%. Since the method does not run the PSI-BLAST method, it takes <<1 sec to process each sequence.

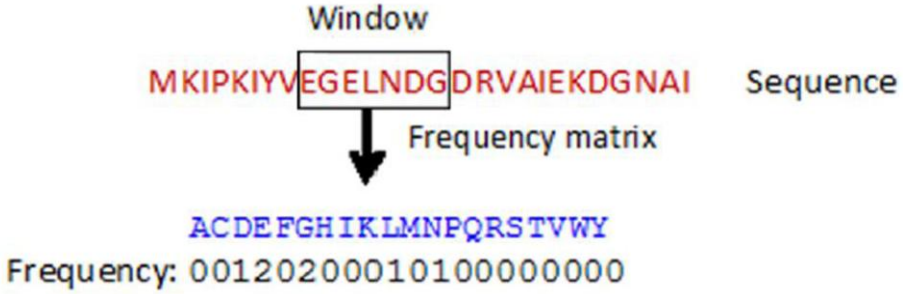

Figure S1. Illustration of amino acid frequency counting within a window size of three residues.

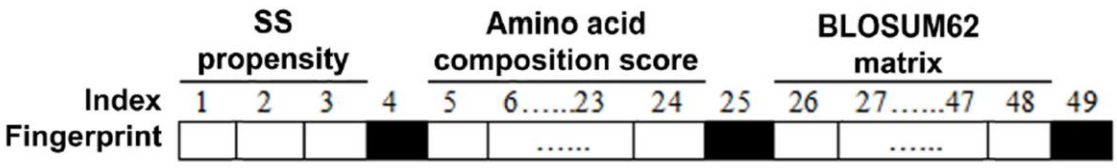

Figure S2. Illustration of fingerprint assignments from neural network SS training where SS propensity score, amino acid composition score, and BLOSUM62 substitution matrix are listed side-by-side with a separation of a noise (black filled cell).

Solvent accessibility prediction. The solvent accessibility (SA) of an amino acid is categorized into three states based on the relative solvent accessibility (RSA) of the amino acid in the protein structure. An amino acid is defined as Buried (B) if RSA<0.09, or Exposed (E) if RSA>0.64, or intermediate (I) otherwise. RSA is calculated by the ratio of the actual accessibility of the amino acid in structures versus that of the amino acid in a tri-peptide conformation (A-X-A) [8].

To predict solvent accessibility, we first computed the SA propensity score of residues in the B, E and I states based on the statistics of the 45,397 non-redundant PDB structures using an equation similar to Eq. S1. The definition of amino acid composition score is same as used in SS prediction but we used 150 as normalization factor. Secondary

structure information computed from our prediction method was incorporated as a binary matrix to improve the SA prediction accuracy. We assume a three state (helix, sheet and otherwise) representation of secondary structure. Therefore, the binary matrix contains three columns corresponding to each state. If an amino acid is present in a particular state the matrix value is assigned as 1, it is 0 otherwise.

The fingerprint scoring matrices in the SS training includes SA propensity (Index 1-3 of Fig S3), secondary structure prediction (Index 5-7 of Fig S3), amino acid composition score (Index 9-28 of Fig S3), and BLOSUM62 substitution matrix (Index 30-52 of Fig S3), which have been listed side-by-side and separated by noises (cells filled with black) in Figure S3.

The one-layer back propagation neural network was trained on the 5,527 non-redundant protein structures with a window size of 12. A separate test on 625 proteins shows that the average SA Q3 accuracy is 66.1%.

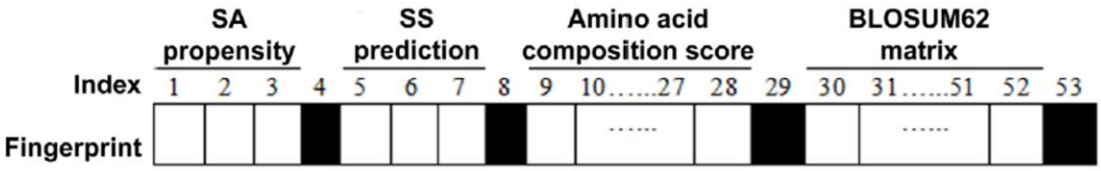

Figure S3. Illustration of fingerprint assignments from SA neural network training where SA propensity score, secondary structure prediction, amino acid composition score, and BLOSUM62 substitution score are listed side-by-side with a separation of a noise (black filled cell).

Backbone torsional angle prediction. The real values of backbone torsion angles (Φ and Ψ) are predicted using a similar back propagation neural-network approach as developed by Xu and Zhang [10]. The input training features include the secondary structure assignment of the target residue and the PSI-BLAST check point file, where the outputs are the real value of the torsion angles. The NN was trained on the same set of 5,527 non-homologous proteins, with true secondary structures and torsional angles were assigned by DSSP [9].

For single-sequence based torsion angle prediction, the input features of the neural network include the secondary structure as predicted by our single-sequence based NN method and the check point file that was converted from BLOSUM 62 substitute matrix. In the validation data, the method shows optimal performance with a window size of 21 for phi angle and 17 for psi angle prediction, which are selected in our predictors. A test on the same set used in the solvent accessibility benchmark of 625 non-homologous proteins, which are non-redundant to the training set, indicates that the average deviation of Φ and Ψ from the DSSP assigned experimental values are comparable to ANGLOR [4], which was trained on more computationally expensive PSI-BLAST PSSM profiles, with the deviation in Φ being only 4.3° higher than ANGLOR and Ψ only 1.5° higher.

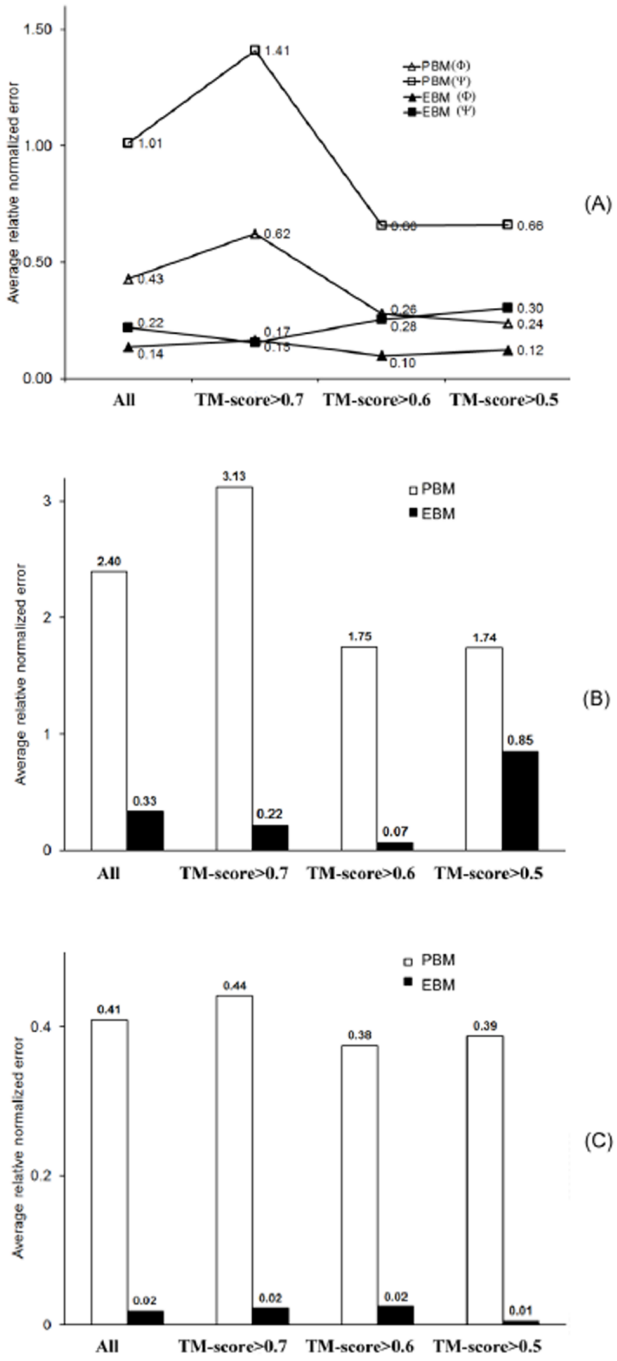

Figure S4. The average normalized relative error (NRE) of the structural features of the designed sequence relative to the DSSP assignments. (A) Backbone torsion angles (Φ/Ψ); (B) Secondary structure (SS); (C) Solvent accessibility (SA). Along the X-axis, the dataset is divided based on the TM-score cutoff on the templates that were used for constructing sequence profiles.

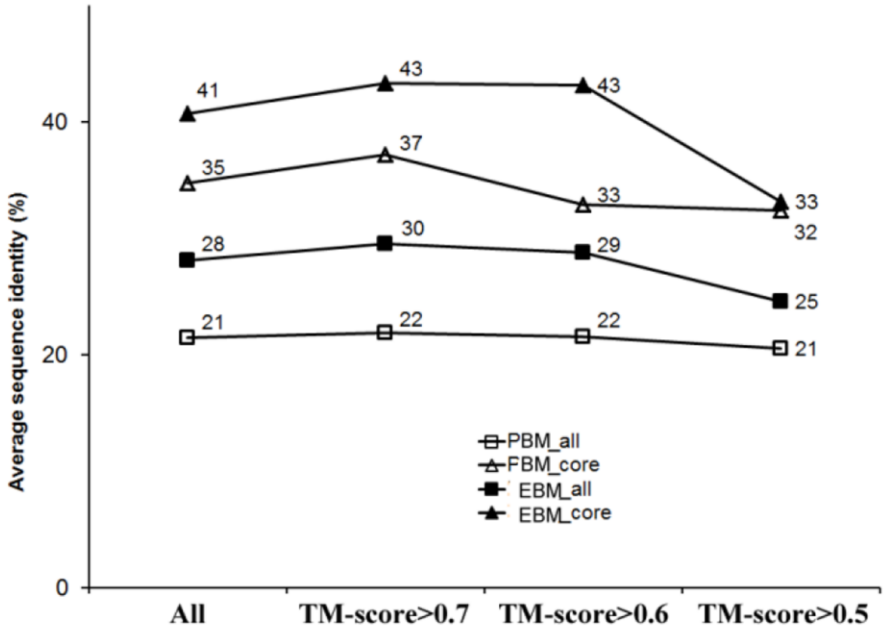

Figure S5. Average sequence identity of the designed sequences to the target sequences. 'All' indicates overall sequence identity and 'core' indicates the identity at the core of the proteins. Along with X-axis, the dataset is divided based on TM-score cutoff on the template proteins that are used for constructing the sequence profiles.

***DNA and protein sequences*.** The designed DNA sequences were optimized based on frequent codon usage in *E. coli* (K-12 strain). Ligation independent cloning (LIC) handles are in bold. Protein sequence lengths are given below. The N-terminal cloning residues "SNA" (lower case in sequence) remain after rTEV protease cleavage during purification, extending the length of the purified proteins by three amino acids.

DNA sequence of designed hNRNPK domain (original target scaffold, PDB ID:1ZZK):

TACTTCCAATCCAATGCACAGGGTTCTGACATCACCCTGCAGATCTCTATC
CCGACCAACATGATCGGTGCGGTTATCGGTAAAGGTGGTGAAGTTATCAAA
GAAATCCAGGAAAAAACCGGTGCGCGTATCCAGATGTCTAAACCGGAAGGT
GGTGACAAAGAAAAAATGGTTACCGTTACCGGTCCGCCGGAATCTATCGAA
AAAGCGAAAGAACTGATCATCGAAATGGTTGAAGAATCTCAGGGTCAGAAA
TTC**TAACATTGGAAGTGGATAA**

*Protein sequence of designed hnRNPK (*Length 80/83 designed/expressed)

snaQGSDITLQISIPTNMIGAVIGKGGEVIKEIQEKTGARIQMSKPEGGDKEKMVTV
TGPPESIEKAKELIIEMVEESQGQKF

DNA sequence of designed thioredoxin domain (original target scaffold, PDB ID: 1R26):

TACTTCCAATCCAATGCATCTATCGTTAAAGTTCAGTCTCCGGAAAACTTCC
AGGAAATCATCAAAGCGGGTAAACTGGTTGTTATCTACTTCTACGCGCCGTG

GTGCCCGCCGTGCCAGAAAGTTTCTCCGGAAATGGAAGCGATGGCGAAAGAA
TACGAAAACGTTATGTTCATCGCGGTTGACATCAACCACAACGAAGAACTGG
CGAAAAAATTCAACATCCAGGAACTGCCGACCATCCTGATCATCAAAGACGG
TAAAATCATGGCGTCTGTTACCGGTGCGAAACCGGAAGAAGTTTCTGAATAC
ATCTCTCAGCTGCTGCGTGAA**TAACATTGGAAGTGGATAA**

DNA sequence of designed thioredoxin (Length: 105/108 designed/expressed):

snaSIVKVQSPENFQEIIKAGKLVVIYFYAPWCPPCQKVSPEMEAMAKEYENVMF
IAVDINHNEELAKKFNIQELPTILIIKDGKIMASVTGAKPEEVSEYISQLLRE

DNA sequence of designed CISK-PO domain (original target scaffold, PDB ID: 1XTE):

TACTTCCAATCCAATGCACCGGACTCTCTGATGAAAGTTTCTATCCCGGACT
TCGAAAAAGAAGGTGAAGGTAAATCTAAACACGTTATGTACAAAATCAAAGT
TAAAACCGGTGGTGAAGAATGGGCGGTTTACCGTCGTTACTCTGACTTCTACT
GGCTGCACAAAAAACTGCAGCAGCGTTACCCGGAACTGGTTCCGGAACTGCC
GCCGAAAAAATGGATCTACTCTGCGCTGGACGAACAGATCCTGGAAAAACGT
AAACAGGGTCTGGAAAAATACATCCAGCGTATCGTTTCTCACCCGGTTCTGG
CGAACGACGAACTGGTTGTTTCTTTCCTGCAGGCGAAAGCGGAACACACCGG
T**TAACATTGGAAGTGGATAA**

Protein sequence of designed CISK-PX (Length: 116/119 designed/expressed):

snaPDSLMKVSIPDFEKEGEGKSKHVMYKIKVKTGGEEWAVYRRYSDFYWLHKK
LQQRYPELVPELPPKKWIYSALDEQILEKRKQGLEKYIQRIVSHPVLANDELVVSF
LQAKAEHTG

DNA sequence of designed Lov2 domain (original target scaffold, PDB ID:2V0U):

TACTTCCAATCCAATGCATCTTCTGGTCAGTCTCAGAACCTGGAAAACGCGG
AACAGACCTTCATCATCACCGACCCGCGTCTGCCGGACGGTCCGATCGTTTAC
GCGTCTGAAGGTTTCCTGAACCTGACCGGTTACGGTCGTGAAGAAATCCTGG
GTCGTAACTGCCGTTTCCTGCAGGGTCCGGCGACCGACCCGGCGACCGTTCA
GGAAATGCGTAACGCGCTGTCTAACGAAGAACCGTGGACCGTTGAACTGATC
AACTACAAAAAAGACGGTACCAAATTCTGGAACATCCTGACCATGGTTCCGG
TTAAAGACAACGACGGTGAAGTTATGTACTACATCGGTGTTCAGATGGACGT
TACCAAACACCGTAAAGACCGTGCGGAAGACGAAGCGATGATGTACGTTGTT
AAAACCGCGCAGGGTATCATGGAACTGATGAAAGCGATG**TAACATTGGAAG**
TGGATAA

Protein sequence of designed Luv2 (Length: 146/149 designed/expressed):

snaSSGQSQNLENAEQTFIITDPRLPDGPIVYASEGFLNLTGYGREEILGRNCRFLQ
GPATDPATVQEMRNALSNEEPWTVELINYKKDGTKFWNILTMVPVKDNDGEVM
YYIGVQMDVTKHRKDRAEDEAMMYVVKTAQGIMELMKAM

DNA sequence of designed TIF1 domain (original target scaffold, PDB ID:3IO4):

TACTTCCAATCCAATGCATTCAAAGAAATGATCGACGGTATCGTTATCCGTA
CCAACGGTAACGGTATCTTCAAAGTTGAACTGAAAAACGGTATGAAAGTTAT
GTGCCACGTTCGTGACAAAATCAAAGAAAACAAAGCGACCATCAAACCGGG

TGACTACGTTCTGGTTCGTCTGGTTCGTAAAGACCCGGTTCGTGGTACCATCA
TGGGTATCCTGGAA**TAACATTGGAAGTGGATAA**

Protein sequence of designed TIF1 (Length: 68/71 designed/expressed):

snaFKEMIDGIVIRTNGNGIFKVELKNGMKVMCHVRDKIKENKATIKPGDYVLVRLVR
KDPVRGTIMGILE

Table S1. List of designed proteins from previous experiments.

Sequence ID	PDB ID	Length	Failed step[a]	References
1lq7A	1lq7A	67	Success	Dai et al [11]
1mj0A	1mj0A	156	Success	Kohl et al [12]
1p68A	1p68A	102	Success	Wei et al [13]
1qysA	1qysA	91	Success	Kuhlman et al [14]
1vjqA	1vjqA	71	Success	Dantas et al [15]
2a3dA	2a3dA	73	Success	Walsh et al [16]
2cw1A	2cw1A	65	Success	Walsh et al [16]
2juaA	2juaA	102	Success	Go et al [17]
2jvfA	2jvfA	94	Success	Stordeur et al [18]
2kl8A	2kl8A	85	Success	Koga et al [19]
2ln3A	2ln3A	83	Success	Koga et al [19]
2ltaA	2ltaA	110	Success	Koga et al [19]
2lv8A	2lv8A	110	Success	Koga et al [19]
2lvbA	2lvbA	112	Success	Koga et al [19]
2qyjA	2qyjA	154	Success	Merz et al [20]
3b83A	3b83A	95	Success	Hu et al [21]
MSP2-FC27	/	237	Natively unfolded	Zhang et al [22]
alpha-synuclein	/	140	Natively unfolded	Eliezer et al [23]
4EbindingproteinI	/	118	Natively unfolded	Fletcher et al [24]
4EbindingproteinII	/	120	Natively unfolded	Fletcher et al [24]
Fold-I-11	/	76	Not soluble	Koga et al [19]
Fold-I-2	/	76	Not folded-CD	Koga et al [19]
Fold-I-4	/	76	Not folded-CD	Koga et al [19]
Fold-I-8	/	76	Not folded NMR	Koga et al [19]
Fold-II-8	/	99	Not folded-CD	Koga et al [19]
Fold-II-9	/	99	Not folded-CD	Koga et al [19]
Fold-II-1	/	99	Not folded NMR	Koga et al [19]
Fold-II-3	/	99	Not folded NMR	Koga et al [19]
Fold-II-11	/	99	Not folded NMR	Koga et al [19]
Fold-III-6	/	72	Not soluble	Koga et al [19]
Fold-III-7	/	72	Not soluble	Koga et al [19]
Fold-III-1	/	72	Incorrect CD	Koga et al [19]
Fold-III-13	/	72	Not folded-CD	Koga et al [19]
Fold-III-4	/	72	Not folded NMR	Koga et al [19]
Fold-IV-2	/	101	Not folded NMR	Koga et al [19]
Fold-V-8	/	99	Not soluble	Koga et al [19]
Fold-V-1	/	99	Not folded-CD	Koga et al [19]
Fold-V-2	/	99	Not folded-CD	Koga et al [19]
Fold-V-4	/	99	Not folded-CD	Koga et al [19]
Fold-V-9	/	99	Not folded-CD	Koga et al [19]
Fold-V-10	/	99	Not folded-CD	Koga et al [19]
Fold-V-11	/	99	Not folded-CD	Koga et al [19]

Fold-V-12	/	99	Not folded-CD	Koga et al [19]
MDM2a	/	85	Not folded-CD	Shultis et al [25]
MDM2b	/	85	Not folded-CD	Shultis et al [25]

[a]**Success**: protein structure was solvable by NMR or X-ray crystallography; **Natively unfolded**: biologically expressed (i.e. not designed) protein experimentally known to be unfolded in the native state; **Not soluble**: protein was not soluble after expression (usually indicative of improperly exposed hydrophobic surfaces); **Not folded-CD**: largely random coil secondary structure by CD; **Incorrect CD**: secondary structure by CD not consistent with predicted structure; **Not folded NMR**: poorly resolved NMR spectra reflective of a dynamic (unstable) protein core.

Table S2. Summary of EBM design on the 87 test proteins.

PDBID	L[a]	SC[b]	aTM[c]	TM[d]	R (Å)[e]	Normalized relative error				Sequence identity (%)	
						SS	Φ	Ψ	SA	All	Core
1A2P_A	109	a+b	0.36	0.99	0.37	0.06	0.17	0.08	-0.05	42	44
1ABA_A	88	a/b	0.66	0.87	1.93	0.20	0.27	0.17	-0.01	34	54
1BKR_A	109	a	0.82	0.99	0.29	-0.22	0.02	-0.11	-0.02	43	63
1DBW_A	124	a/b	0.76	0.88	1.71	0.00	-0.03	-0.05	-0.08	28	38
1EAQ_A	125	b	0.64	0.91	1.45	-0.10	-0.06	-0.20	-0.13	30	54
1EW4_A	107	a+b	0.50	0.95	0.92	0.92	0.24	0.42	0.05	17	21
1F46_A	140	a+b	0.55	0.98	0.45	-0.09	0.01	0.09	-0.05	28	42
1GBS_A	186	a+b	0.67	0.99	0.18	1.30	0.02	-0.04	0.04	46	53
1GUT_A	52	b	0.74	0.97	0.36	0.33	0.09	-0.05	0.08	25	44
1HZT_A	151	a+b	0.77	0.99	0.45	-0.13	0.43	0.11	0.04	25	33
1I2T_A	62	a	0.59	0.96	0.51	0.14	-0.20	0.58	0.13	20	50
1IDP_A	148	a+b	0.79	0.94	1.20	0.03	-0.06	-0.01	-0.02	24	30
1IUJ_A	98	a+b	0.78	0.78	2.22	0.71	-0.07	-0.13	-0.17	25	36
1JB3_A	128	b	0.60	0.98	0.51	-0.04	0.23	0.26	0.03	25	37
1JF8_A	131	a/b	0.90	0.99	0.44	-0.36	0.03	0.44	0.06	35	50
1KMT_A	139	b	0.66	0.99	0.31	0.07	0.09	0.21	-0.04	27	30
1KNG_A	145	a/b	0.76	0.96	0.88	0.19	0.06	0.18	0.01	30	39
1KQ1_A	61	b	0.81	0.96	0.50	0.29	0.16	0.25	0.01	28	62
1M9Z_A	106	small	0.60	0.96	0.83	0.13	0.11	0.06	-0.01	35	59
1MF7_A	195	a/b	0.83	0.95	1.36	0.08	0.26	0.03	0.02	27	36
1MG4_A	102	a+b	0.50	0.63	3.55	1.83	0.50	0.48	0.24	14	23
1NXM_A	195	b	0.56	0.96	1.13	0.27	0.21	0.27	0.00	23	30
1NZ0_A	110	a+b	0.52	0.84	1.75	0.53	0.29	0.53	0.10	14	13
1O7I_A	116	b	0.81	0.94	1.01	0.00	0.20	-0.10	0.00	30	41
1OAI_A	58	a	0.67	0.85	1.48	-0.80	0.60	1.73	0.04	28	31
1OH0_A	126	a+b	0.78	0.95	1.17	-0.05	0.14	0.10	0.00	27	32
1OK0_A	75	b	0.62	0.60	7.22	0.00	0.00	-0.09	0.07	18	45
1QHQ_A	140	b	0.65	0.37	11.00	0.24	0.17	0.06	0.20	35	38
1R26_A	106	a/b	0.79	0.93	1.01	0.20	0.15	-0.08	0.01	34	51
1R6J_A	83	b	0.82	0.89	1.68	0.00	0.39	0.28	0.03	24	40
1SHU_X	182	a/b	0.84	0.91	1.73	0.25	0.25	0.10	0.06	23	38
1T3Y_A	132	a+b	0.88	0.85	1.97	0.00	-0.03	-0.12	0.06	19	33
1TQG_A	106	a	0.75	0.97	0.64	1.50	0.22	1.01	-0.02	19	33
1TUK_A	68	a	0.53	0.38	5.30	0.33	0.12	0.98	0.06	9	11
1UCS_A	65	b	0.75	0.97	0.50	-0.05	-0.13	-0.12	0.05	53	67
1URR_A	98	a+b	0.85	0.95	0.80	0.13	0.08	0.16	-0.01	34	50
1UTG_A	71	a	0.50	0.98	0.42	2.33	-0.25	0.30	-0.19	10	10
1V5I_B	71	a+b	0.65	0.87	1.36	-0.43	0.10	0.77	0.02	24	45
1VH5_A	139	a+b	0.81	0.98	0.64	0.88	0.45	0.27	0.03	25	33
1VKK_A	138	a+b	0.86	0.91	1.70	0.06	0.19	0.38	0.00	29	32
1VQS_A	105	a+b	0.66	0.98	0.56	0.05	0.49	0.28	0.03	15	21

1VZI_A	126	b	0.32	0.18	16.93	1.93	0.36	0.79	0.30	17	33
1WLU_A	118	a+b	0.79	0.98	0.45	0.14	0.33	0.10	0.00	29	26
1X6Z_A	120	a+b	0.70	0.93	1.20	-0.24	0.19	0.13	-0.02	34	43
1XTE_A	117	a+b	0.81	0.97	0.71	0.00	0.27	0.21	-0.05	32	56
1ZHV_A	129	a+b	0.80	0.57	8.26	1.08	0.26	1.14	0.18	18	21
1ZKE_A	82	a	0.61	0.92	1.11	-0.22	-0.16	-0.47	0.05	14	18
1ZZK_A	81	a+b	0.79	0.89	2.99	-0.29	0.57	0.19	0.00	36	50
2ANX_A	146	a+b	0.81	0.99	0.30	0.04	0.18	0.04	0.06	30	34
2BWF_A	76	a+b	0.81	0.90	1.31	-0.35	0.34	0.15	0.00	32	65
2C9Q_A	103	b	0.63	0.86	1.83	0.22	-0.18	0.17	0.02	30	36
2CAR_A	197	a/b	0.78	0.99	0.44	-0.06	0.05	0.04	-0.49	29	36
2CMP_A	53	a	0.52	0.94	0.61	3.00	0.00	0.63	0.21	25	46
2CVI_A	78	a+b	0.76	0.89	1.17	0.08	0.14	0.21	-0.04	30	41
2D3D_A	84	a	0.66	0.91	1.15	-0.16	0.22	0.39	-0.05	31	54
2ERB_A	124	a	0.78	0.99	0.37	0.33	0.09	0.09	0.01	29	47
2F01_A	122	b	0.70	0.98	0.76	0.04	-0.07	-0.18	-0.09	39	59
2FTR_A	97	a+b	0.79	0.90	1.36	0.00	0.01	0.00	-0.06	26	36
2GMY_A	148	a	0.64	0.99	0.29	0.42	0.33	0.45	0.02	20	30
2GPI_A	92	a+b	0.49	0.82	1.92	0.43	0.19	0.11	0.07	9	17
2J2J_A	183	b	0.90	0.77	3.62	-0.35	0.23	0.14	0.10	23	36
2J5Y_A	55	a	0.77	0.91	1.24	4.00	-0.39	-0.04	-0.18	35	55
2J8B_A	79	small	0.70	0.90	1.16	0.00	0.04	0.15	-0.03	32	52
2O1Q_A	140	b	0.68	0.53	10.01	1.08	0.35	0.42	0.22	17	26
2O9S_A	68	b	0.76	0.83	3.15	0.12	0.08	0.43	0.10	39	74
2P5K_A	64	a	0.64	0.96	0.53	0.00	0.02	0.52	-0.01	38	63
2PR7_A	138	a/b	0.78	0.70	4.83	0.25	0.22	0.22	0.25	27	39
2PTH_A	194	a/b	0.71	0.98	0.58	0.00	0.18	0.01	0.00	25	37
2PV2_A	104	a+b	0.91	0.96	0.69	0.36	0.15	-0.11	0.03	32	59
2QCP_X	77	b	0.77	0.94	0.77	0.83	1.14	0.95	0.30	18	42
2V0U_A	147	a/b	0.77	0.92	2.65	0.15	0.02	0.12	-0.03	54	70
2V1Q_A	60	b	0.82	0.91	1.01	0.00	0.22	0.48	0.12	34	36
2VMH_A	148	b	0.71	0.97	0.95	0.18	-0.03	0.22	0.05	24	25
2VPB_A	58	a+b	0.68	0.86	1.15	-0.07	-0.15	-0.13	-0.03	39	64
2VZC_A	128	a	0.79	0.93	1.24	0.29	0.26	0.60	0.18	19	34
2WLV_A	145	a/b	0.75	0.92	1.49	0.05	0.05	0.19	-0.02	33	42
2ZXY_A	87	a	0.63	0.69	4.22	-0.18	0.09	0.02	0.01	36	55
3CTG_A	109	a/b	0.89	0.97	0.76	-0.25	0.05	0.03	0.00	39	54
3E9T_A	113	b	0.86	0.99	0.31	0.11	-0.01	0.03	0.01	39	41
3EBT_A	132	a+b	0.78	0.92	1.47	-0.11	-0.05	0.01	-0.01	29	33
3EF8_A	150	a+b	0.74	0.80	8.05	0.52	0.39	0.29	-0.02	23	27
3FEA_A	84	a+b	0.57	0.97	0.57	-0.06	0.18	-0.06	-0.02	42	69
3FIL_A	57	a+b	0.84	0.92	0.73	-0.11	-0.20	-0.20	0.01	63	92
3G21_A	78	a	0.45	0.36	11.57	3.57	0.12	1.26	0.18	12	16
3G36_A	53	a	0.43	0.32	11.19	1.67	-0.23	-0.36	-0.45	21	0
3IV4_A	107	a/b	0.80	0.82	1.95	-0.36	-0.07	-0.07	0.04	22	34
3VUB_A	102	b	0.70	0.88	1.54	0.28	0.12	0.33	0.06	18	31
Average	111		0.71	0.87	2.12	0.33	0.14	0.22	0.02	28	41

[a]Length of targets.
[b]SCOP type of targets.
[c]Average TM-score between the first I-TASSER model of design sequence and the structural analogs used for profiling.
[d]TM-score computed between the first I-TASSER model of design sequence and the target scaffold.
[e]RMSD computed between the first I-TASSER model of design sequence and the target scaffold.

Table S3. Summary of EvBM design on the 87 test proteins.

PDBID	TM-score[a]	RMSD (Å)[b]	Normalized relative error				Sequence identity (%)	
			SS	Φ	Ψ	SA	All	Core
1A2P_A	0.99	0.35	0.12	0.31	0.03	-0.03	48	44
1ABA_A	0.69	3.41	3.20	0.64	0.23	0.30	18	25
1BKR_A	0.99	0.28	0.33	0.09	-0.10	-0.05	42	55
1DBW_A	0.85	2.02	-0.08	0.14	0.00	0.00	29	44
1EAQ_A	0.92	1.42	-0.26	-0.03	-0.19	-0.11	30	37
1EW4_A	0.94	0.92	0.75	0.61	0.34	-0.07	16	31
1F46_A	0.97	0.66	0.27	0.25	-0.23	-0.01	27	29
1GBS_A	0.99	0.16	1.35	0.19	0.07	0.13	54	65
1GUT_A	0.91	0.77	-0.33	0.09	-0.12	0.03	24	11
1HZT_A	0.95	1.08	0.00	0.42	0.25	0.04	21	23
1I2T_A	0.93	0.76	0.29	0.12	0.69	0.20	25	57
1IDP_A	0.75	3.69	0.16	0.17	0.08	0.14	26	21
1IUJ_A	0.72	3.13	0.57	0.06	0.60	-0.19	27	43
1JB3_A	0.48	6.07	0.12	0.32	-0.10	0.28	26	33
1JF8_A	0.78	3.04	0.09	0.31	0.93	0.26	28	34
1KMT_A	0.95	1.04	0.14	0.10	0.01	-0.02	28	30
1KNG_A	0.96	0.95	0.50	0.17	0.11	0.03	26	34
1KQ1_A	0.81	1.65	0.29	0.79	0.44	0.08	23	54
1M9Z_A	0.92	1.26	0.19	0.22	0.68	0.01	45	51
1MF7_A	0.96	1.16	0.14	0.29	0.15	0.07	37	57
1MG4_A	0.45	8.43	1.67	0.45	0.52	0.14	13	20
1NXM_A	0.94	1.48	0.09	0.33	0.36	0.11	26	24
1NZ0_A	0.58	6.80	0.74	0.24	0.37	0.23	14	25
1O7I_A	0.99	0.39	-0.15	0.32	-0.17	0.02	22	33
1OAI_A	0.65	5.06	0.80	1.00	1.25	0.14	21	8
1OH0_A	0.96	0.95	0.05	0.27	0.00	0.04	31	35
1OK0_A	0.62	4.06	-0.14	0.05	-0.11	0.03	14	25
1QHQ_A	0.51	9.60	0.31	0.32	0.34	-0.07	30	40
1R26_A	0.90	1.33	0.20	0.21	-0.07	0.07	30	31
1R6J_A	0.91	1.48	-0.25	0.32	-0.06	0.07	32	50
1SHU_X	0.96	1.21	0.13	0.40	-0.05	-0.04	25	42
1T3Y_A	0.73	3.57	0.08	0.23	0.31	0.15	19	24
1TQG_A	0.94	0.96	1.25	0.04	0.16	-0.02	24	21
1TUK_A	0.25	11.13	0.44	0.52	2.10	0.02	10	16
1UCS_A	0.85	1.86	-0.16	-0.01	-0.12	-0.09	50	72
1URR_A	0.94	0.96	0.13	0.11	0.73	0.00	31	38
1UTG_A	0.83	2.08	3.33	-0.19	0.41	-0.29	10	10
1V5I_B	0.63	3.25	-0.29	0.34	0.03	-0.15	10	0
1VH5_A	0.98	0.48	0.13	0.41	0.22	0.06	31	28
1VKK_A	0.93	1.45	0.13	0.24	1.07	0.02	26	28
1VQS_A	0.68	3.31	-0.11	0.41	0.33	0.03	20	7
1VZI_A	0.34	11.67	0.52	0.43	-0.01	0.23	6	8
1WLU_A	0.98	0.49	0.00	0.31	0.08	0.03	32	41

1X6Z_A	0.67	4.23	0.53	0.52	0.42	-0.03	31	22
1XTE_A	0.89	2.06	0.27	0.25	0.42	-0.09	26	44
1ZHV_A	0.65	5.86	0.56	0.54	1.05	0.26	14	18
1ZKE_A	0.26	14.51	0.06	-0.03	0.01	0.14	12	18
1ZZK_A	0.83	3.11	0.29	0.48	0.16	0.03	41	64
2ANX_A	0.99	0.30	0.09	0.08	0.30	0.00	28	34
2BWF_A	0.92	1.21	-0.35	0.34	0.27	-0.02	40	65
2C9Q_A	0.77	2.51	0.39	0.28	0.16	-0.29	19	29
2CAR_A	0.96	1.06	-0.03	0.12	0.13	0.06	31	40
2CMP_A	0.61	2.27	4.00	0.34	0.36	0.00	17	23
2CVI_A	0.92	0.95	-0.17	0.32	0.58	0.03	29	36
2D3D_A	0.90	1.29	0.16	0.39	1.24	0.04	24	38
2ERB_A	0.99	0.35	0.50	0.18	0.27	0.01	32	39
2F01_A	0.98	0.75	-0.22	-0.08	-0.05	0.12	43	52
2FTR_A	0.73	2.73	0.44	0.62	0.42	0.05	27	24
2GMY_A	0.97	1.44	0.42	0.32	0.35	0.20	23	26
2GPI_A	0.46	8.26	0.93	0.42	0.42	-0.13	12	17
2J2J_A	0.78	3.54	-0.19	0.23	0.20	-0.14	21	31
2J5Y_A	0.93	0.86	3.00	-0.31	0.32	-0.04	33	36
2J8B_A	0.91	1.23	3.33	0.14	0.21	0.19	41	52
2O1Q_A	0.64	9.04	0.71	0.49	0.32	0.03	19	24
2O9S_A	0.81	3.53	0.25	0.25	-0.07	0.03	34	53
2P5K_A	0.93	0.73	0.33	0.17	1.12	0.28	30	38
2PR7_A	0.70	5.72	0.22	0.37	0.36	0.03	19	25
2PTH_A	0.96	1.37	0.29	0.29	-0.10	0.04	36	51
2PV2_A	0.95	0.84	0.36	0.10	0.19	0.23	33	53
2QCP_X	0.89	1.16	1.67	1.15	0.89	0.07	14	21
2V0U_A	0.93	2.74	0.05	0.10	0.24	0.07	59	68
2V1Q_A	0.88	1.11	0.20	0.25	0.33	0.05	29	29
2VMH_A	0.90	2.13	0.14	0.18	0.11	0.02	24	22
2VPB_A	0.83	1.27	0.07	-0.18	0.44	0.13	32	55
2VZC_A	0.91	1.48	0.24	0.43	0.79	0.15	20	34
2WLV_A	0.91	1.62	-0.11	0.10	0.25	-0.02	42	49
2ZXY_A	0.81	1.98	0.09	0.45	0.06	-0.04	33	41
3CTG_A	0.97	0.74	-0.12	0.15	-0.09	-0.02	43	51
3E9T_A	0.99	0.29	0.33	0.08	0.10	0.03	41	50
3EBT_A	0.89	1.63	0.00	0.18	0.15	0.04	29	28
3EF8_A	0.80	7.14	0.28	0.31	0.29	-0.06	28	33
3FEA_A	0.97	0.54	-0.12	0.45	0.24	0.04	35	58
3FIL_A	0.92	0.75	-0.11	-0.01	-0.21	0.01	52	75
3G21_A	0.29	15.25	2.71	0.15	1.27	-0.32	12	16
3G36_A	0.41	5.03	3.67	0.18	0.84	0.02	13	0
3IV4_A	0.81	2.11	-0.36	0.04	0.04	0.00	20	24
3VUB_A	0.71	3.10	0.38	0.20	0.11	0.11	20	31
Average	0.82	2.82	0.48	0.26	0.30	0.04	27	35

[a]TM-score between the first I-TASSER model of design sequence and the target scaffold.
[b]RMSD between the first I-TASSER model and the target scaffold.

Table S4. Summary of PBM design on the 87 test proteins.

PDBID	TM-score[a]	RMSD (Å)[b]	Normalized relative error				Sequence Identity (%)	
			SS	Φ	Ψ	SA	All	Core
1A2P_A	0.91	1.45	2.88	0.23	1.11	0.73	21	38
1ABA_A	0.94	0.91	0.80	-0.07	0.01	0.05	33	58
1BKR_A	0.98	0.45	2.22	0.09	0.21	0.48	24	40
1DBW_A	0.90	1.49	3.17	0.44	0.92	0.34	20	27
1EAQ_A	0.89	1.53	0.46	-0.08	0.24	-0.04	23	37
1EW4_A	0.26	11.88	2.58	0.45	1.04	0.30	12	7
1F46_A	0.98	0.52	0.18	0.07	0.07	0.15	23	33
1GBS_A	0.87	2.42	3.00	0.56	0.30	0.65	19	27
1GUT_A	0.76	1.49	3.17	-0.01	0.67	0.08	20	44
1HZT_A	0.97	0.79	1.17	1.03	2.58	0.54	23	33
1I2T_A	0.74	1.79	0.29	-0.14	0.23	0.33	15	21
1IDP_A	0.20	16.66	2.13	0.54	1.70	0.61	17	30
1IUJ_A	0.63	4.21	2.86	0.26	2.92	0.81	19	21
1JB3_A	0.98	0.57	0.92	0.06	0.11	0.07	25	39
1JF8_A	0.99	0.36	2.18	0.46	1.40	0.42	25	44
1KMT_A	0.29	10.24	3.03	0.75	2.11	0.45	21	32
1KNG_A	0.99	0.27	1.88	0.69	1.55	0.43	24	28
1KQ1_A	0.21	14.84	5.57	1.27	2.47	0.34	17	31
1M9Z_A	0.18	15.92	3.38	0.64	2.02	0.46	8	8
1MF7_A	0.92	1.68	1.89	0.74	1.16	0.70	14	20
1MG4_A	0.84	2.28	3.50	0.34	0.44	0.40	24	37
1NXM_A	0.89	2.56	1.79	0.47	1.20	0.62	18	19
1NZ0_A	0.66	4.18	1.16	0.05	0.16	0.74	23	38
1O7I_A	0.98	0.48	2.77	0.88	1.16	0.35	26	44
1OAI_A	0.85	1.37	1.00	1.09	2.00	0.37	18	23
1OH0_A	0.95	1.12	2.47	1.52	3.17	0.65	24	24
1OK0_A	0.72	2.27	0.32	-0.12	0.22	0.21	27	55
1QHQ_A	0.98	0.47	0.76	0.18	0.86	0.27	21	33
1R26_A	0.75	2.78	9.80	1.36	3.33	0.47	16	23
1R6J_A	0.88	1.51	3.75	0.30	1.27	0.44	32	65
1SHU_X	0.96	1.17	0.25	0.09	-0.14	0.04	21	24
1T3Y_A	0.19	15.86	1.65	0.58	0.73	0.54	17	33
1TQG_A	0.41	6.43	1.50	0.53	0.76	0.83	13	18
1TUK_A	0.77	1.75	-0.11	0.01	0.17	-0.04	25	42
1UCS_A	0.96	0.81	0.05	0.13	0.75	0.45	30	61
1URR_A	0.96	0.87	4.25	0.67	2.70	0.49	23	44
1UTG_A	0.92	0.86	1.00	-0.21	0.48	0.22	17	30

1V5I_B	0.64	3.66	1.36	0.44	0.95	0.36	26	50
1VH5_A	0.96	0.87	7.00	0.83	2.00	0.26	20	30
1VKK_A	0.21	17.15	3.13	0.63	2.81	0.57	16	28
1VQS_A	0.70	5.75	1.05	0.58	1.66	0.48	23	25
1VZI_A	0.97	0.69	1.74	0.40	0.94	0.40	22	42
1WLU_A	0.97	0.75	9.00	2.12	4.41	0.40	25	38
1X6Z_A	0.91	1.47	1.65	0.27	0.37	0.27	24	27
1XTE_A	0.85	2.90	3.09	0.30	1.50	0.37	17	19
1ZHV_A	0.98	0.64	1.40	0.46	1.56	0.38	29	49
1ZKE_A	0.90	1.95	-0.39	-0.27	-0.61	0.72	22	29
1ZZK_A	0.84	2.85	1.29	0.20	0.02	0.59	26	45
2ANX_A	0.88	3.04	0.83	0.43	0.56	0.74	21	40
2BWF_A	0.87	1.35	0.65	0.71	2.77	0.26	17	35
2C9Q_A	0.78	2.56	2.26	0.34	1.55	0.62	25	39
2CAR_A	0.99	0.32	-0.03	-0.08	-0.03	-0.61	25	39
2CMP_A	0.82	1.27	5.50	0.42	1.42	0.54	25	69
2CVI_A	0.34	9.84	3.08	1.37	3.11	0.58	23	50
2D3D_A	0.90	1.12	0.26	0.26	0.46	0.87	25	54
2ERB_A	0.87	2.84	3.33	0.20	0.76	0.67	19	22
2F01_A	0.44	6.69	1.89	0.51	1.18	0.45	19	21
2FTR_A	0.32	9.23	3.00	0.42	0.83	0.39	23	28
2GMY_A	0.95	1.17	2.42	0.72	1.39	0.58	15	26
2GPI_A	0.25	13.64	2.50	0.35	0.95	0.78	16	35
2J2J_A	0.17	17.86	1.65	0.09	0.84	0.14	20	34
2J5Y_A	0.37	11.98	11.00	0.04	-0.07	0.43	13	36
2J8B_A	0.20	9.20	11.33	0.49	0.88	0.41	9	12
2O1Q_A	0.97	0.92	0.17	0.09	0.71	-0.14	28	33
2O9S_A	0.80	3.24	3.62	0.31	1.33	0.61	34	63
2P5K_A	0.25	11.87	5.33	1.11	3.81	0.38	19	38
2PR7_A	0.98	0.61	-0.44	-0.32	-0.11	-0.05	30	33
2PTH_A	0.94	1.37	0.52	0.18	0.12	0.11	22	29
2PV2_A	0.97	0.66	2.09	1.04	1.22	0.67	24	44
2QCP_X	0.95	0.74	4.33	0.74	1.09	0.34	26	63
2V0U_A	0.93	4.26	-0.50	0.21	-0.04	0.10	17	26
2V1Q_A	0.86	1.13	5.60	0.88	1.81	0.51	27	64
2VMH_A	0.77	3.10	2.95	0.27	1.29	0.64	23	31
2VPB_A	0.21	9.99	0.43	-0.09	0.64	0.41	11	18
2VZC_A	0.99	0.27	0.06	-0.01	0.07	-0.05	28	43
2WLV_A	0.97	0.83	0.89	0.05	0.30	0.42	20	27
2ZXY_A	0.20	12.48	2.27	0.04	-0.26	0.64	20	27
3CTG_A	0.97	0.69	0.81	0.73	0.82	0.36	25	35

3E9T_A	0.98	0.60	6.67	0.43	1.67	0.57	22	41
3EBT_A	0.67	4.54	3.44	1.84	2.66	0.70	27	47
3EF8_A	0.22	18.02	2.76	0.93	1.97	0.48	17	29
3FEA_A	0.97	0.51	0.94	0.45	0.66	0.62	34	58
3FIL_A	0.89	0.93	1.89	-0.03	0.30	0.49	25	58
3G21_A	0.36	8.68	2.00	0.10	0.42	0.77	16	5
3G36_A	0.52	4.50	1.00	-0.14	0.52	0.03	15	0
3IV4_A	0.79	2.46	1.00	0.39	0.75	0.17	14	26
3VUB_A	0.91	1.17	1.03	0.13	0.47	0.24	19	34
Average	0.74	4.14	2.40	0.43	1.11	0.41	21	35

[a]TM-score between the first I-TASSER model of design sequence and the target scaffold.
[b]RMSD between the first I-TASSER model and the target scaffold.

Table S5. Summary of EBM based protein design on the *M. tuberculosis* genome. The data has been sorted by the normalized relative error (NRE) on secondary structure.

PDBID	L[a]	Scp[b]	Np[c]	SI%[d]	R(Å)[e]	Normalized relative error				Sequence identity (%)	
						SS	Φ	Ψ	SA	All	Core
3FAV_A	74		167	9	2.14	-0.47	-0.54	-0.64	-0.47	9	17
3FAV_B	72		34	10	2.34	-0.40	-0.32	0.08	0.03	17	18
2G38_A	77	a	60	10	2.46	-0.36	-0.09	-0.25	-0.71	18	58
3H6P_C	56		20	13	3.23	-0.33	0.12	0.13	-0.01	14	20
1IM5_A	179	a/b	22	25	0.28	-0.29	0.03	0.01	-0.02	39	59
2ZHX_B	82	a+b	338	7	0.89	-0.28	0.11	0.21	-0.05	17	20
1P3H_A	99	b	53	17	1.06	-0.27	0.07	-0.07	0.06	26	47
3KNH_H	138		10	41	1.89	-0.26	0.06	-0.12	0.00	46	44
3KNI_S	98		27	15	3.57	-0.25	-0.07	-0.08	0.23	17	27
2PMU_A	98		11	30	1.44	-0.25	0.05	0.46	0.03	36	50
1UVH_A	157	a	88	13	2.42	-0.23	-0.11	-0.03	-0.05	22	28
3MPZ_A	125		11	38	0.42	-0.20	0.23	0.39	0.08	50	55
3NFW_A	179		20	20	0.28	-0.19	-0.05	0.02	-0.02	30	51
3M6C_A	194		30	16	0.87	-0.19	0.01	0.10	0.07	31	44
2W72_B	146		71	25	0.89	-0.19	-0.16	-0.09	0.00	30	38
3IOS_A	134		93	18	2.02	-0.18	0.07	0.10	0.03	28	39
2CGQ_A	74		31	23	1.24	-0.18	-0.23	-0.01	-0.06	35	62
1TFU_A	157	a/b	28	22	0.91	-0.16	-0.03	-0.06	0.01	36	47
2A6P_A	193		22	23	2.65	-0.15	0.01	-0.08	0.17	30	41
2H7V_A	178	a/b	95	24	1.43	-0.15	0.05	0.36	-0.01	33	38
2NM3_A	121	a+b	13	20	0.33	-0.14	0.02	-0.01	0.03	31	50
3RIH_A	260		224	25	0.85	-0.14	0.06	-0.11	0.02	33	42
3HX9_A	101		16	16	1.15	-0.14	0.21	0.41	-0.09	27	31
2Z2I_A	179	a+b	60	12	1.75	-0.13	0.21	0.20	0.03	16	19
3NY4_A	265		48	24	1.45	-0.13	0.13	0.10	0.01	26	38
3P0T_A	136		10	21	1.00	-0.12	0.06	0.06	-0.03	30	55
3KNI_D	271		10	36	1.89	-0.12	0.09	0.14	-0.05	37	47
3NBK_A	158		30	25	0.91	-0.12	-0.01	0.05	-0.02	39	53
3P2J_A	191		44	15	1.22	-0.12	0.01	0.27	-0.01	27	34
2Z90_A	161		72	13	6.67	-0.12	-0.12	0.08	-0.07	17	21
2VKL_A	77		22	12	3.26	-0.11	0.22	-0.12	0.18	12	33
1U2P_A	156		12	27	1.09	-0.10	0.13	-0.12	0.01	37	52
3MYB_A	259		66	23	2.03	-0.09	0.04	0.05	0.04	28	44
3KNH_E	150		10	27	1.28	-0.09	0.13	0.12	-0.01	35	48
2YW6_A	148	a	98	14	1.52	-0.09	-0.13	-0.21	-0.11	20	20
3A2G_A	154		72	22	0.95	-0.08	0.26	0.08	-0.02	34	37
2QXX_A	189		12	19	1.02	-0.08	0.30	0.13	0.04	32	45
2XWL_A	221		50	18	0.42	-0.08	0.24	-0.07	-0.01	23	32
3MXU_A	134		11	28	4.36	-0.08	-0.07	-0.09	0.17	35	47
3QXZ_A	259		66	23	1.89	-0.07	0.06	0.13	0.09	32	47
3KNH_G	155		10	40	3.80	-0.07	0.27	0.14	0.20	37	48

3KNH_I	127		14	28	1.07	-0.07	-0.05	-0.01	0.00	38	38
3R9T_A	260		70	26	18.09	-0.06	0.04	-0.03	0.00	31	50
1J7S_B	146	a	71	26	0.94	-0.06	-0.05	-0.10	0.00	36	46
2VVR_A	147		11	41	0.86	-0.06	-0.12	0.03	0.01	44	57
2C45_A	113		66	13	0.52	-0.06	-0.09	-0.19	-0.01	24	34
3MOY_A	258		67	27	1.15	-0.06	0.04	0.10	0.03	34	51
2A15_A	133	a+b	47	15	0.91	-0.05	0.09	0.48	-0.07	22	15
1ZNW_A	182	a/b	12	31	0.55	-0.05	0.08	0.04	-0.04	37	55
3KNI_6	50		22	19	5.20	-0.05	-0.04	-0.13	-0.06	12	20
3C4I_A	99		15	26	0.51	-0.05	-0.13	-0.09	-0.01	37	60
3HGB_A	134		10	43	0.35	-0.05	-0.02	-0.07	-0.01	50	65
3KNI_Y	100		10	27	4.79	-0.04	-0.10	0.08	0.04	34	56
3H81_A	256		69	28	1.04	-0.04	0.17	-0.04	-0.01	36	55
3HH2_A	109		10	34	0.60	-0.03	0.12	-0.17	0.07	39	78
1YK9_A	184		13	28	7.79	-0.03	0.08	0.29	0.35	24	39
3KNI_E	204		11	30	10.20	-0.03	0.21	0.13	0.63	15	20
2QZ8_A	146		10	32	1.34	-0.03	0.02	-0.11	0.02	36	53
2NTV_A	268		188	18	2.85	-0.03	0.19	0.15	0.00	24	27
1P9L_A	245	a/b	107	12	0.96	-0.03	0.43	0.31	0.04	19	26
3FGN_A	230		13	16	0.50	-0.03	0.32	-0.02	0.00	20	31
3GWK_C	98		17	17	1.29	0.00	-0.23	0.10	-0.17	30	54
3KNH_J	98		10	35	1.24	0.00	-0.12	-0.14	0.04	40	55
2X89_A	128		105	37	2.35	0.00	-0.04	-0.17	0.03	52	58
3KNI_R	117		20	17	1.52	0.00	-0.08	-0.11	-0.01	29	41
3EUL_A	123		178	18	0.74	0.00	0.00	-0.11	0.00	33	45
3KNI_W	113		10	37	0.89	0.00	-0.09	0.13	-0.01	43	56
3C57_A	51		19	28	6.53	0.00	-0.24	-0.31	0.06	29	63
2C2X_A	280		48	16	1.16	0.00	0.02	0.00	0.04	26	35
2W29_A	147		10	32	0.92	0.00	-0.04	0.04	0.03	37	47
3HUG_A	80		22	13	3.00	0.00	0.40	0.12	0.11	26	75
1L4U_A	165	a/b	61	17	3.02	0.00	0.01	-0.04	-0.02	21	29
3LHD_A	253		21	19	0.80	0.00	0.33	0.40	-0.01	25	30
3R20_A	217		10	27	0.17	0.00	0.29	-0.10	0.04	36	52
1W66_A	218	a+b	14	14	0.25	0.00	0.09	0.03	0.05	23	30
2O8X_A	61		16	22	1.66	0.00	0.19	-0.01	-0.19	30	45
3KNI_2	71		31	17	0.86	0.00	0.04	-0.28	0.02	23	70
1WQG_A	184	a+b	10	24	1.45	0.00	0.20	-0.01	0.00	40	48
1XZ0_B	98	b	119	20	1.96	0.00	0.04	-0.06	0.31	29	50
2W6X_A	154		72	22	0.93	0.00	0.06	0.14	0.03	31	38
1W19_A	147		10	36	0.59	0.00	0.06	0.12	0.00	40	54
3KNI_X	92		10	35	1.60	0.00	0.06	-0.13	-0.03	36	48
3IJF_X	123		10	35	0.82	0.00	0.21	0.38	-0.05	40	50
3D55_A	68		19	12	1.40	0.00	0.17	0.34	0.26	16	17
2I1U_A	108		137	24	2.52	0.00	0.27	0.52	0.00	38	65
1NXJ_A	156	a/b	30	16	0.22	0.00	0.10	0.30	0.04	35	49
1DF7_A	159	a/b	32	27	0.27	0.00	0.16	0.05	0.04	34	46
2FVH_A	99		86	11	0.93	0.00	0.22	0.23	-0.01	23	39
3ICO_A	244		15	29	0.30	0.03	0.01	0.08	0.04	31	46
2W3B_A	185		29	27	0.82	0.03	0.11	0.20	0.01	38	51

3NDO_A	225		252	12	2.58	0.03	0.07	0.02	0.01	23	29
3QKA_A	233		71	25	2.24	0.03	0.19	-0.07	0.06	33	47
2W25_A	147		10	25	1.24	0.03	-0.01	-0.01	0.04	29	41
1U5H_A	223	a/b	68	13	0.45	0.03	0.11	-0.03	0.01	23	29
2XB9_A	158		202	10	0.98	0.04	-0.02	0.32	0.09	20	32
3KNI_U	117		14	24	1.27	0.04	-0.05	-0.09	-0.01	38	56
1Y25_A	165	a/b	42	22	1.03	0.04	0.23	0.00	0.09	29	38
1NBU_A	118	a+b	18	23	0.76	0.05	-0.12	-0.07	-0.05	33	59
1NBU_C	118	a+b	16	19	0.72	0.05	-0.02	-0.04	0.00	32	49
3KNH_N	60		10	21	2.36	0.05	-0.04	0.01	0.10	33	57
3GWM_A	129		13	27	3.41	0.05	0.35	0.18	0.07	34	50
3KNH_B	234		27	14	1.84	0.05	0.00	-0.05	-0.02	19	31
2JFR_A	234		14	22	3.52	0.05	0.29	0.37	0.09	22	34
1J7Y_B	146	a	71	25	0.85	0.06	-0.06	-0.17	0.00	30	38
3I86_A	136		10	38	1.02	0.06	-0.05	0.02	0.06	49	62
3KNI_V	101		12	12	4.53	0.07	-0.03	0.14	0.54	22	43
3KNH_F	101		95	10	4.27	0.07	0.19	0.71	-0.03	21	41
2Y88_A	244		46	15	4.99	0.07	0.25	0.05	0.04	23	28
2CDN_A	186	a/b	26	28	1.22	0.07	0.13	0.29	0.04	36	53
3I7T_A	118		26	26	1.58	0.07	-0.06	-0.03	0.08	36	60
3NE3_B	129		12	29	2.30	0.07	0.26	-0.09	0.04	33	46
1NAZ_A	154	a	71	22	1.12	0.08	0.39	0.23	0.07	24	29
3PK0_A	260		223	25	0.63	0.08	0.02	0.09	0.02	26	37
3QHA_A	277		52	16	14.76	0.08	0.19	-0.05	0.05	24	38
2A11_A	154		10	24	0.93	0.08	0.11	-0.24	-0.03	37	51
3ASE_A	154		71	22	0.89	0.08	0.17	0.08	0.02	34	43
2IVM_A	147		10	33	1.01	0.09	-0.02	-0.03	0.04	39	56
3ONR_A	71		14	14	2.44	0.09	0.24	0.13	0.05	27	78
3KNI_F	207		10	30	5.10	0.09	0.03	0.18	0.02	28	31
3HZA_A	162		11	32	0.27	0.10	0.07	0.22	-0.01	41	64
3SBX_A	177		11	27	0.36	0.10	-0.05	0.45	0.00	31	39
2CHC_A	169	a+b	42	12	4.03	0.10	0.37	0.19	-0.04	24	33
3PL1_A	185		20	25	5.31	0.10	0.19	0.20	0.27	22	32
3QJ7_A	260		17	47	0.82	0.10	0.11	0.04	0.08	41	49
3ODG_A	273		25	23	1.57	0.11	0.22	0.15	0.08	34	40
2V2X_A	275	b	19	36	7.19	0.11	0.34	0.18	0.37	25	29
3HNT_L	214		17	41	3.60	0.11	0.13	0.05	0.16	38	55
3L60_A	229		24	15	1.06	0.11	0.15	0.24	0.03	21	30
3KNI_O	122		10	56	3.65	0.11	0.15	0.01	0.28	50	55
1J7W_A	141	a	74	27	0.83	0.13	0.08	0.05	-0.07	38	53
1Y0H_A	101	a+b	36	15	1.19	0.13	0.04	0.28	0.13	18	23
2WT9_A	207		17	23	0.32	0.13	0.19	0.05	0.03	29	44
3KNH_P	83		10	19	0.84	0.13	0.05	0.06	0.02	24	30
2R5U_A	144		20	13	0.40	0.13	0.27	0.13	-0.01	28	48
3F8L_A	162		14	18	0.38	0.14	0.29	0.10	0.05	21	28
1D0I_A	150	a/b	495	10	1.07	0.15	0.16	0.00	0.11	19	30
3HSS_A	274		79	16	0.69	0.15	0.09	0.08	0.00	21	31
2VE6_B	99	b	120	19	1.63	0.15	0.03	0.16	0.02	33	55
2WUG_A	283		75	19	1.32	0.16	0.27	0.26	0.03	26	33

3R2N_A	119		11	32	0.48	0.17	0.26	0.39	0.03	45	52
3QK8_A	259		70	22	17.19	0.17	0.18	0.39	0.05	25	33
3I4O_A	68		21	20	1.67	0.17	-0.09	0.11	-0.17	32	58
1X89_A	174	b	19	23	2.46	0.18	0.04	0.42	0.04	28	52
3KNI_Q	141		10	24	2.64	0.19	0.00	0.03	-0.03	40	61
1L1E_A	272	a/b	33	16	1.77	0.19	0.01	0.11	0.02	24	27
3H5J_A	168		12	25	1.26	0.19	-0.12	-0.16	0.01	36	47
1GTV_A	208	a/b	17	20	2.38	0.20	0.44	0.10	0.03	27	41
1LU4_A	134	a/b	86	16	1.93	0.20	0.28	0.31	0.03	23	40
3K3C_A	152		73	10	2.55	0.20	0.16	0.06	0.06	22	29
2VE6_A	276	a+b	19	35	10.65	0.21	0.35	0.27	0.41	25	33
1TPY_A	285	a/b	22	18	1.14	0.21	0.01	0.05	0.03	24	29
2BM5_A	183	b	21	12	4.28	0.21	0.15	0.22	0.01	17	26
3TDT_A	274	b	10	22	3.09	0.21	0.14	-0.02	0.10	36	44
2GLN_A	128		85	14	1.03	0.22	0.09	0.18	-0.04	32	50
2QRW_A	126		76	11	0.65	0.23	0.29	-0.12	0.15	18	30
3KNI_N	138		11	39	1.12	0.24	-0.02	0.09	0.01	41	60
3F8M_A	230		16	16	10.65	0.25	0.15	0.35	0.05	23	26
2V2X_B	100	b	113	20	1.61	0.25	-0.01	-0.16	0.10	29	44
2B70_A	162	a+b	13	88	4.53	0.25	0.39	0.60	0.34	46	41
3NE0_A	208		10	33	0.65	0.26	0.11	0.10	0.08	50	69
3PPI_A	258		221	22	1.04	0.26	0.11	0.16	0.03	29	40
3NB5_A	257		21	44	3.42	0.26	0.24	0.20	0.15	39	50
3DWG_C	92		13	19	1.29	0.27	-0.19	-0.23	0.01	37	50
2G4R_A	153		14	28	1.14	0.27	-0.09	0.07	-0.01	33	43
1PQW_A	183	a/b	62	20	2.88	0.27	0.13	-0.05	0.06	27	35
2O03_A	129		10	16	9.12	0.29	0.16	-0.02	0.08	27	52
3RFQ_A	154		14	28	0.23	0.32	0.16	0.29	0.01	38	51
3QUA_A	179		11	28	2.64	0.33	0.23	0.49	0.12	31	39
3OCC_A	237		37	23	1.49	0.33	0.16	0.08	-0.01	27	29
3PO8_A	98		30	23	1.12	0.33	0.10	0.09	0.03	29	39
3KNI_P	146		10	14	4.51	0.33	0.12	0.06	0.20	21	20
2XW7_A	175		30	15	4.62	0.34	0.51	0.29	0.15	20	29
1N3N_B	99	b	118	19	2.70	0.37	0.04	0.43	0.20	25	37
3G1M_A	194		48	16	1.43	0.38	0.07	0.08	-0.01	18	21
2XPG_A	274		19	22	12.09	0.38	0.19	0.38	0.37	29	34
2QVB_A	296		59	17	5.08	0.39	0.02	-0.03	0.27	15	19
3OC6_A	243		15	29	4.42	0.40	0.06	0.41	0.27	21	32
3AJX_A	207		632	12	0.49	0.43	0.34	0.05	0.02	19	24
1QPO_A	284	a/b	15	32	3.22	0.43	0.31	0.41	0.34	29	36
3D0S_A	224		10	31	4.74	0.45	0.24	0.54	0.06	27	38
1UZR_A	282	a	20	21	2.84	0.46	0.01	0.32	0.37	18	26
3KNI_H	163		10	32	7.20	0.50	0.00	0.05	0.26	26	29
2VG0_A	227		44	9	1.74	0.52	0.12	0.29	0.12	21	25
2WZM_A	274		33	32	4.80	0.53	0.31	0.28	0.19	24	34
3R0O_A	262		68	25	18.90	0.54	0.34	0.38	0.18	24	32
2VVP_A	157	a/b	10	27	3.62	0.55	0.53	0.57	0.01	32	42
1DT0_A	197	a+b	34	45	4.75	0.55	0.29	0.32	0.47	31	39
2IB0_A	142		98	11	3.88	0.56	0.16	0.29	0.01	14	25

1W74_A	171	b	30	37	3.63	0.56	0.26	0.13	0.15	33	46
2FR2_A	161	b	18	11	6.16	0.57	0.54	1.06	0.05	22	25
3KNH_T	99		74	10	5.09	0.57	-0.04	0.11	0.03	19	56
3K6Y_A	227		10	20	4.93	0.58	0.19	0.33	0.17	24	33
3G5O_B	81		63	10	0.88	0.58	-0.05	0.56	0.12	16	26
1G2O_A	262	a/b	29	19	5.06	0.59	0.63	0.51	0.33	21	27
2W72_A	141		74	27	3.29	0.62	0.29	0.33	0.33	25	38
2VZW_A	146		33	15	1.49	0.64	0.25	0.51	0.03	20	33
2WWO_A	158		19	33	5.56	0.68	0.29	0.42	0.22	32	45
1I9G_A	264	a/b	15	20	3.18	0.68	0.76	0.80	0.30	16	23
1W5R_A	273	a+b	10	44	5.38	0.70	0.39	0.55	0.28	30	34
2H7M_A	268	a/b	189	19	6.82	0.71	0.61	0.54	0.05	15	16
1R88_A	267	a/b	42	11	5.30	0.73	0.26	0.13	0.44	12	18
3HUG_B	62		161	8	1.33	0.75	0.10	0.02	0.19	19	14
2VLR_E	240	b	23	33	10.14	0.76	0.41	0.32	0.23	30	34
2G2D_A	160		10	36	2.40	0.77	0.19	-0.06	0.09	36	51
1NGK_A	126	a	77	12	3.94	0.79	0.16	0.03	0.10	22	32
3R9Q_A	233		72	26	5.16	0.79	0.35	0.29	0.15	19	27
3PKA_A	284		26	25	11.45	0.79	0.43	0.57	0.34	25	32
1PM4_A	117	b	88	7	11.91	0.81	-0.06	0.34	0.17	13	21
2O7G_A	88		12	13	0.43	0.83	0.24	0.45	0.00	23	44
2W3G_A	150		26	17	0.50	0.85	0.45	0.33	0.06	27	36
1NFF_A	244	a/b	253	25	4.88	0.85	0.37	0.33	0.34	24	34
3HST_B	139		10	25	7.38	0.85	0.28	1.03	0.29	24	33
1PZS_A	171	b	19	24	8.92	0.87	0.47	0.52	0.53	19	28
1YLK_A	163		11	22	5.22	0.88	0.10	0.41	0.06	24	30
2ZFZ_A	77		36	13	1.00	0.89	0.08	0.21	0.08	23	39
3KNI_G	181		11	35	8.91	0.90	0.21	0.34	0.41	30	44
2Q7Y_B	98	a+b	119	19	3.07	0.92	0.13	0.26	0.17	26	50
3G5O_A	92		12	9	1.42	0.92	0.18	1.09	-0.06	21	50
3KNH_K	119		38	12	1.29	1.00	-0.10	0.08	-0.04	24	35
3PVV_A	96		68	9	0.36	1.00	-0.20	0.07	-0.07	20	32
2VLR_D	199	b	19	56	8.36	1.04	0.35	0.26	0.40	20	24
2ZYQ_A	296		15	22	6.35	1.10	0.50	0.63	0.35	22	26
3QD8_A	172		27	19	5.17	1.11	0.21	0.25	0.32	14	22
1K44_A	135	a+b	29	48	6.30	1.17	0.41	0.61	0.34	39	52
1LMI_A	131	b	73	7	2.02	1.17	0.31	0.55	0.18	17	17
1XZ0_A	271	b	17	30	17.14	1.19	0.38	0.68	0.37	20	28
3F9R_A	246		24	18	8.43	1.19	0.22	0.32	0.27	20	27
1S8N_A	190	a/b	227	12	9.40	1.24	0.36	1.26	0.20	25	35
3KNH_S	78		56	12	1.97	1.26	-0.14	-0.05	0.16	26	32
3H87_A	136		33	11	0.60	1.26	0.60	1.16	0.07	25	47
3HNT_H	220		28	43	15.89	1.30	0.56	0.45	0.66	17	20
1F0N_A	284	a/b	13	15	9.75	1.38	0.73	0.39	0.40	15	21
3KNI_7	48		56	13	5.79	1.40	-0.14	-0.06	0.15	23	33
2AQ6_A	143	b	16	15	5.06	1.40	0.52	0.98	0.36	13	11
3OL3_A	98		100	8	10.66	1.42	0.06	0.72	-0.12	15	37
1YK3_A	198	a+b	43	11	3.38	1.50	0.28	0.33	0.25	17	18
3MAY_A	86		213	8	2.48	1.83	-0.12	-0.13	0.09	20	31

1IDS_A	198	a+b	35	41	11.17	1.85	0.39	0.85	0.54	25	32
3E3U_A	196		13	27	12.14	1.91	0.52	0.26	0.54	22	32
3HC7_A	252		56	9	17.46	2.03	0.30	0.42	0.27	19	27
2IG3_A	127		78	9	5.04	2.67	0.33	0.52	0.32	15	23
3PGX_A	270		210	24	12.42	2.69	0.73	1.01	0.41	25	36
2C92_A	147		10	36	6.59	2.83	0.40	0.43	0.51	22	29
Average[f]	133		46	22	2.57	0.27	0.10	0.16	0.07	29	41
Average[g]	169		51	23	3.28	0.29	0.14	0.18	0.09	28	39

[a]Length of protein target.
[b]SCOP type provided whenever they are available in SCOP database.
[c]The number of structural analogs used to construct the structural profile.
[d]The average sequence identity between the target sequence and the sequence of the structural analogs used for profiling.
[e]RMSD between the first I-TASSER model and the target scaffold.
[f]Average results on the cases with length <200 AA.
[g]Average results on all cases.

References

1. Guerois R, Nielsen JE, Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol 320: 369-387.
2. Faraggi E, Zhang T, Yang Y, Kurgan L, Zhou Y (2012) SPINE X: improving protein secondary structure prediction by multistep learning coupled with prediction of solvent accessible surface area and backbone torsion angles. J Comput Chem 33: 259-267.
3. Chen H, Zhou HX (2005) Prediction of solvent accessibility and sites of deleterious mutations from protein sequence. Nucleic Acids Res 33: 3193-3199.
4. Wu S, Zhang Y (2008) ANGLOR: a composite machine-learning algorithm for protein backbone torsion angle prediction. PLoS ONE 3: e3400.
5. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323: 533-536.
6. Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A 89: 10915-10919.
7. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. Nucleic Acids Res 28: 235-242.
8. Rost B, Sander C (1994) Conservation and prediction of solvent accessibility in protein families. Proteins 20: 216-226.
9. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22: 2577-2637.
10. Xu D, Zhang Y (2012) Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins 80: 1715-1735.
11. Dai QH, Tommos C, Fuentes EJ, Blomberg MR, Dutton PL, et al. (2002) Structure of a de novo designed protein model of radical enzymes. J Am Chem Soc 124: 10952-10953.
12. Kohl A, Binz HK, Forrer P, Stumpp MT, Pluckthun A, et al. (2003) Designed to be stable: crystal structure of a consensus ankyrin repeat protein. Proc Natl Acad Sci U S A 100: 1700-1705.
13. Wei Y, Kim S, Fela D, Baum J, Hecht MH (2003) Solution structure of a de novo protein from a designed combinatorial library. Proc Natl Acad Sci U S A 100: 13270-13273.
14. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, et al. (2003) Design of a novel globular protein fold with atomic-level accuracy. Science 302: 1364-1368.
15. Dantas G, Corrent C, Reichow SL, Havranek JJ, Eletr ZM, et al. (2007) High-resolution structural and thermodynamic analysis of extreme stabilization of human procarboxypeptidase by computational protein design. Journal of Molecular Biology 366: 1209-1221.
16. Walsh ST, Cheng H, Bryson JW, Roder H, DeGrado WF (1999) Solution structure and dynamics of a de novo designed three-helix bundle protein. Proc Natl Acad Sci U S A 96: 5486-5491.

17. Go A, Kim S, Baum J, Hecht MH (2008) Structure and dynamics of de novo proteins from a designed superfamily of 4-helix bundles. Protein Sci 17: 821-832.
18. Stordeur C, Dalluge R, Birkenmeier O, Wienk H, Rudolph R, et al. (2008) The NMR solution structure of the artificial protein M7 matches the computationally designed model. Proteins 72: 1104-1107.
19. Koga N, Tatsumi-Koga R, Liu G, Xiao R, Acton TB, et al. (2012) Principles for designing ideal protein structures. Nature 491: 222-227.
20. Merz T, Wetzel SK, Firbank S, Pluckthun A, Grutter MG, et al. (2008) Stabilizing ionic interactions in a full-consensus ankyrin repeat protein. Journal of Molecular Biology 376: 232-240.
21. Hu X, Wang H, Ke H, Kuhlman B (2008) Computer-based redesign of a beta sandwich protein suggests that extensive negative design is not required for de novo beta sheet design. Structure 16: 1799-1805.
22. Zhang X, Perugini MA, Yao S, Adda CG, Murphy VJ, et al. (2008) Solution conformation, backbone dynamics and lipid interactions of the intrinsically unstructured malaria surface protein MSP2. Journal of Molecular Biology 379: 105-121.
23. Eliezer D, Kutluay E, Bussell R, Jr., Browne G (2001) Conformational properties of alpha-synuclein in its free and lipid-associated states. Journal of Molecular Biology 307: 1061-1073.
24. Fletcher CM, McGuire AM, Gingras AC, Li H, Matsuo H, et al. (1998) 4E binding proteins inhibit the translation factor eIF4E without folded structure. Biochemistry 37: 9-15.
25. Shultis D, Mitra P, Zhang Y (2013) Unpublished results.