Supporting Information

Table of Contents

Supporting Texts

Text S1. Fragments generated by the L-BFGS system used in "Zhang_Ab_initio" (D-QUARK-FM)

Apart from the "QUARK" server, the "Zhang_Ab_Initio" server implemented fully *ab initio* structure modeling (D-QUARK-FM) where LOMETS3/gapless threading fragments were not used in any part of the pipeline. In "Zhang_Ab_Initio", the distance and orientation restraints were used to construct rough full-length conformations by L-BFGS gradient descent with the following notation for orientations (t_{ij}) and distances (d_{ij}):

$$E(t_{ij}) = -\log\left(\frac{P(t_{ij}) + \varepsilon}{\varepsilon}\right) \tag{S1}$$

$$E(d_{ij}) = -\log\left(\frac{P(d_{ij}) + \varepsilon}{P(d_{cut}) + \varepsilon}\right) + 1.57 \cdot \log\left(\frac{d_{ij}}{d_{cut}}\right) \tag{S2}$$

Where $d_{cut} = 20\text{Å}$, and $\varepsilon = 1.0 \times \text{E-04}$. Using different cutoffs ranging from 0.55 to 0.95 for the probability of $d_{ij} > 20$ Å, 30 gradient descent runs were performed to generate 30 initial conformations. From these 30 conformations, continuous fragments ranging from one to 20 residues were extracted to build the position-specific fragments library. Here, the orientation energy potential (**Eq. S1**) was also used in the D-QUARK folding simulations.

Text S2. Definitions of mean absolute distance errors (*MAEs*)

To assess the accuracy of DeepPotential predicted distances, we define the measure ***MAE_n*** as the mean absolute distance error between the top k*L* (*L* is the protein length) predicted distances and the corresponding distances calculated from the experimentally solved structures. The equation is as follows:

$$MAE_n = \frac{1}{kL}\sum_{(i,j)}^{kL}|d_{i,j}^{pred} - d_{i,j}^{exp}| \tag{S3}$$

where $d_{i,j}^{exp}$ is the Cα-Cα (or Cβ-Cβ) distance between residue *i* and *j* in the experimental structure, and $d_{i,j}^{pred}$ is the predicted Cα-Cα (or Cβ-Cβ) distance between residue *i* and *j* from DeepPotential. Since DeepPotential predicts the probability distribution for each residue pair *i* and *j*, the distance distributions were first ranked by their peak probability (only distances <20Å were considered). Then, the top k*L*-ranked distance distributions were used to calculate the MAE_n, where $d_{i,j}^{pred}$ was estimated as the middle value of the bin where the peak probability was located in. In particular, we used the top 5*L*-ranked long-range (|*i-j*|>23) Cα-Cα distances from DeepPotential to calculate MAE_n since we found it had the maximal Pearson correlation coefficient (PCC) with TM-scores from the predicted models.

Another measure ***MAE_m*** is defined to qualify how well the predicted models fit with the predicted distances from DeepPotential. The equation for MAE_m is defined as:

$$MAE_m = \frac{1}{kL}\sum_{(i,j)}^{kL}|d_{i,j}^{mod} - d_{i,j}^{pred}| \tag{S4}$$

Similar with MAE_n, the top 5*L*-ranked long-range (|*i-j*|>23) Cα-Cα distances from DeepPotential were used to calculate the MAE_m. $d_{i,j}^{mod}$ is the Cα-Cα distance between residues *i* and *j* in the predicted model structure.

Text S3. The normalized number of effective sequences (*Neff*) in an MSA

The depth of a multiple sequence alignment (MSA) was measured by the normalized number of effective sequences (*Neff*) in this work:

$$Neff = \frac{1}{\sqrt{L}}\sum_{i=1}^{N}\frac{1}{1+\sum_{j=1,j\neq i}^{N}I[S_{i,j}\geq 0.8]} \tag{S5}$$

where L is the length of a protein, N is the number of sequences in the MSA, and $S_{i,j}$ is the sequence identity between the i-th and j-th sequences. $I[S_{i,j} \geq 0.8]$ equals 1 if $S_{i,j} \geq 0.8$, or zero otherwise. Therefore, *Neff* is essentially equal to the number of non-redundant sequences (sequence identity<0.8) in the MSA normalized by the protein length.

Supporting Figures
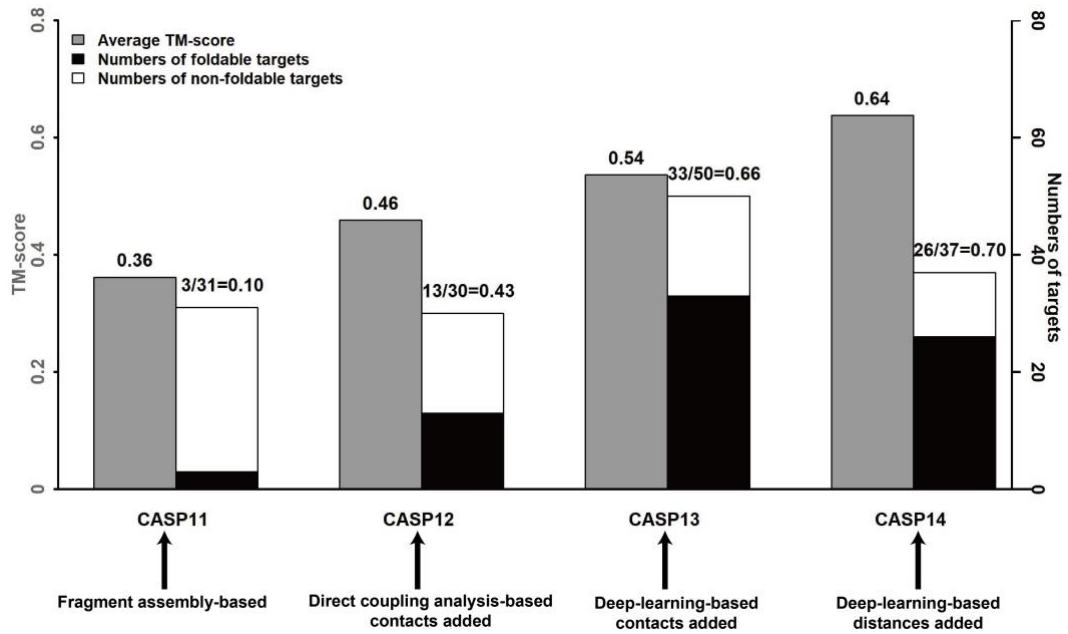
Figure S1. The folding performance for FM targets based on the best models that were submitted by "Zhang-Server" or "QUARK" servers from CASP11 to CASP14. The gray bars represent the average TM-score of the best models from "Zhang-Server" or "QUARK", the black bars indicate the number of foldable targets (TM-score ≥ 0.5), and the white bars depicts the number of non-foldable targets.
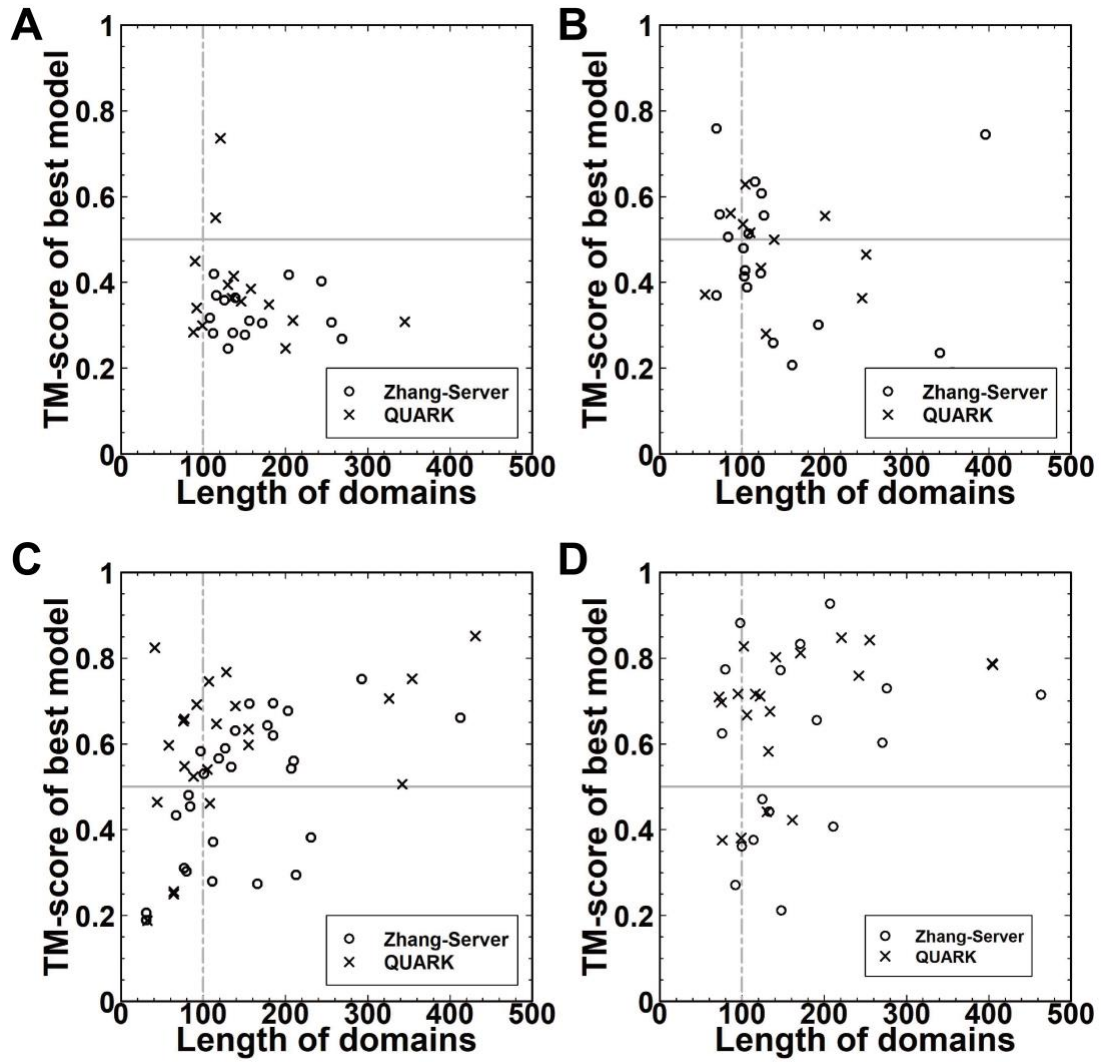
Figure S2. The TM-scores of the best "Zhang-Server" or "QUARK" models for the FM targets in CASP11 (**A**), CASP12 (**B**), CASP13 (**C**) and CASP14 (**D**) vs the domain lengths.
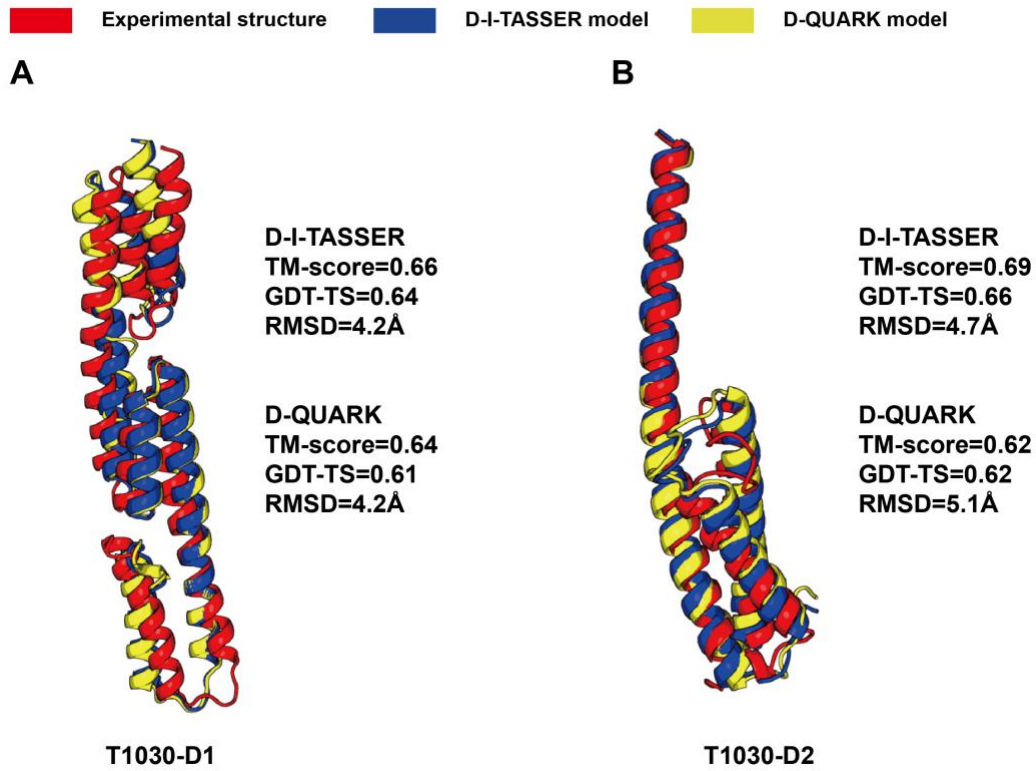
Figure S3. Re-modeling results for T1030-D1 (**A**) and T1030-D2 (**B**) by D-I-TASSER and D-QUARK using the "correct" DeepMSA2 generation. Only the templates that were released before CASP14 were used for the modeling.
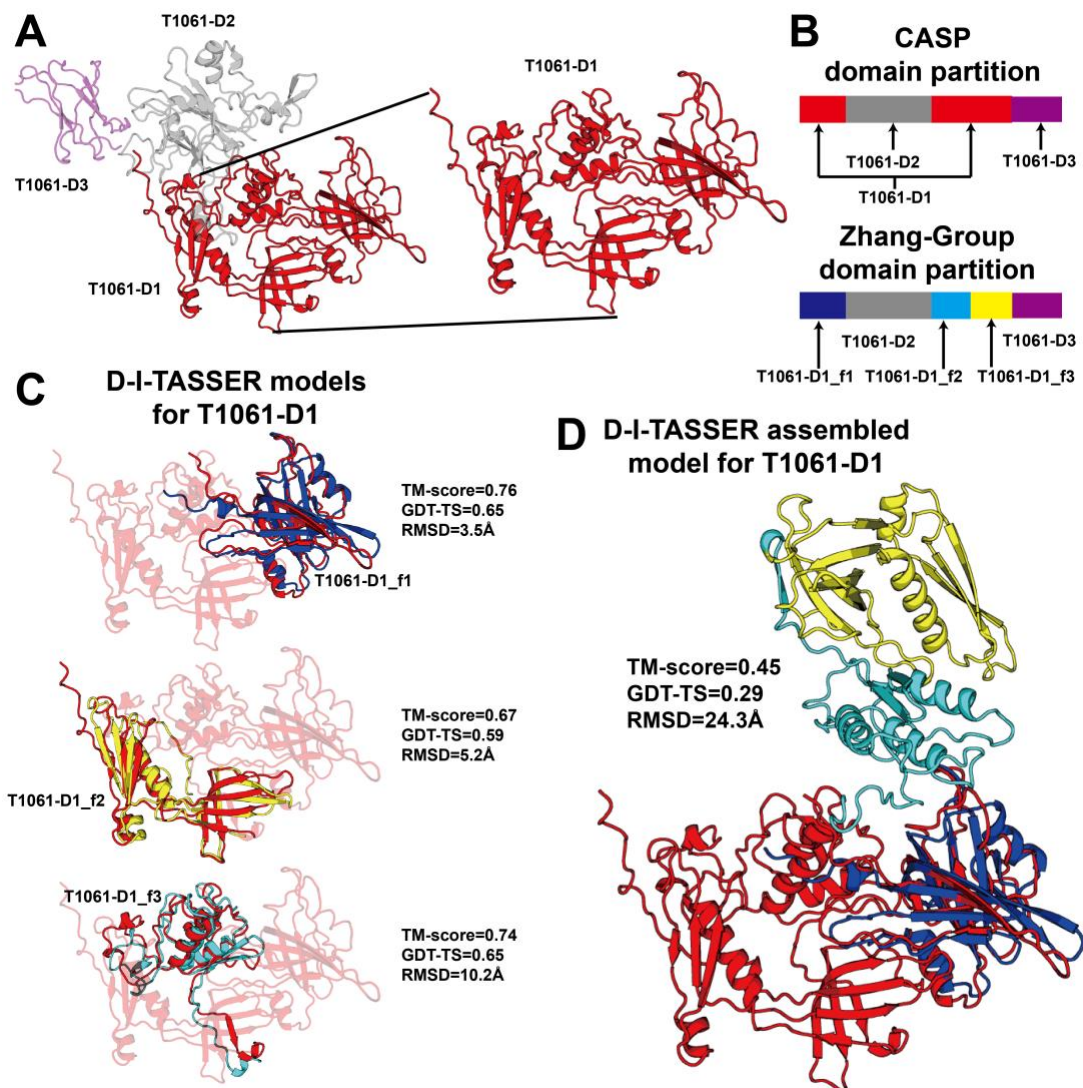
Figure S4. An illustration of modeling information for T1061. (A) Experimental structure for T1061. (B) The domain partition by the CASP assessors and Zhang-Group for T1061. (C) The D-I-TASSER models of three fragments "T1061-D1_f1", "T1061-D1_f2" and "T1061-D1_f3" from T1061-D1 superposed with the experimental structures. (D) The D-I-TASSER model for domain target T1061-D1 superposed with the experimental structures.
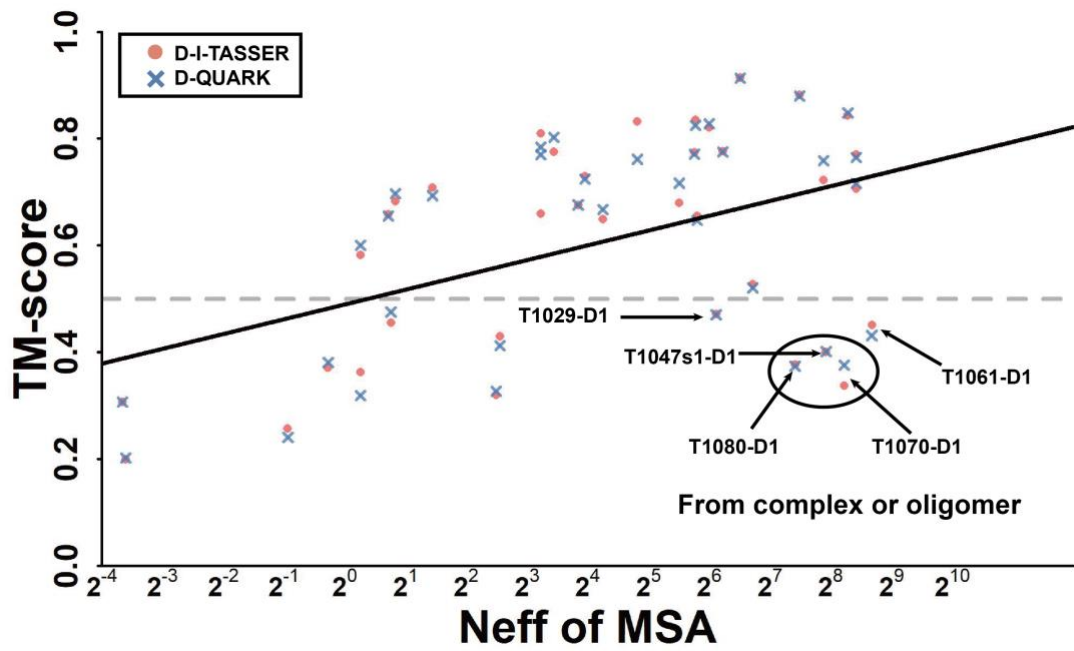
Figure S5. The TM-scores of the D-I-TASSER (red points) and D-QUARK (blue cross) models for the 37 FM targets in CASP14 vs the *Neff* value of the MSAs generated by DeepMSA2.
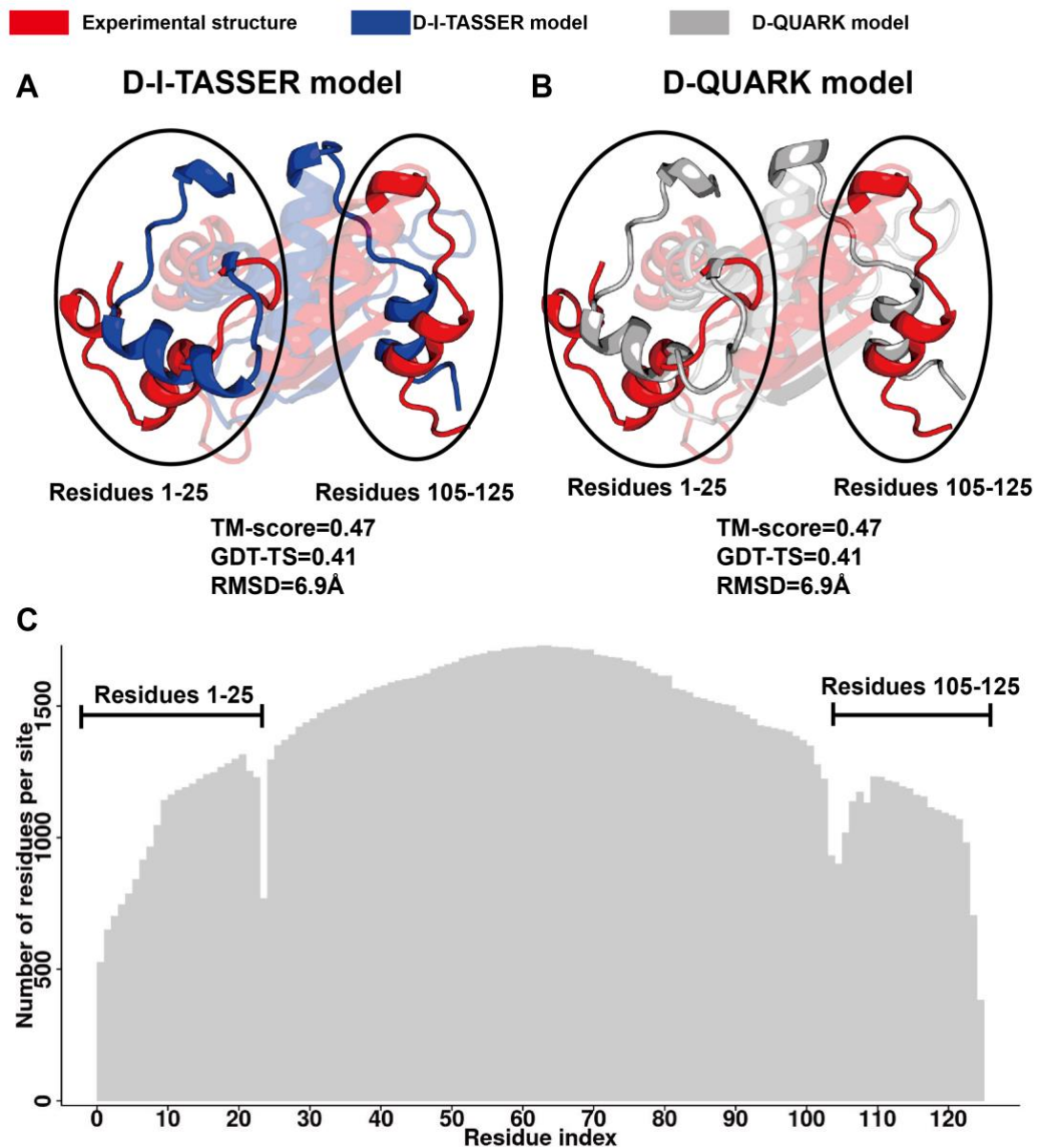
Figure S6. Modeling information for T1029-D1. (**A**) D-I-TASSER model for T1029-D1. The terminal region has been highlighted with the black circles. (**B**) D-QUARK model for T1029-D1. (**C**) The number of effective residues in each sequence position of the MSA generated by DeepMSA2. The two terminal regions (residues 1-25 and residues 105-125) clearly had a smaller number of effective residues than the central region (residues 126-104).

Supporting Tables

Table S1. Performance of Zhang group servers based on average TM-score of the first models and best among five models for 91 CASP14 domain targets. The "Type" column indicates the category of each target. The fourth and sixth columns represent the number of targets with TM-scores greater than 0.5 from the first or the best models. Here, "Zhang-Server" and "QUARK" were based on the D-I-TASSER and D-QUARK algorithms, respectively.

Server	Type	TM-score of first model	#{TM-score$_{first}$>0.5}	TM-score of best model	#{TM-score$_{best}$>0.5}
Zhang-Server	All	0.7065	75	0.7249	78
	TBM	0.7757	51	0.7891	52
	FM	0.6055	24	0.6312	26
QUARK	All	0.7039	74	0.7256	78
	TBM	0.7694	50	0.7901	52
	FM	0.6084	24	0.6314	26
Zhang-CEthreader	All	0.6893	77	0.7194	79
	TBM	0.7600	52	0.7845	52
	FM	0.5862	25	0.6243	27
Zhang-TBM	All	0.6852	75	0.7159	78
	TBM	0.7617	51	0.7846	52
	FM	0.5736	24	0.6158	26
Zhang_Ab_Initio	All	0.6637	71	0.6823	74
	TBM	0.7326	48	0.7523	50
	FM	0.5632	23	0.5802	24

Table S2. Comparison between the first models built by D-I-TASSER, C-I-TASSER and I-TASSER for the 91 CASP14 domain targets. The "Type" column indicates the category of each target. P-values were calculated between the TM-scores of the D-I-TASSER models and the model from the other two methods using one-sided Student's t-tests to test if D-I-TASSER had significantly higher TM-scores. "#{TM-score>0.5}" represents the number of targets with TM-scores greater than 0.5.

Method	Type	TM-score	P-value	#{TM-score>0.5}
D-I-TASSER	All	0.7065	*	75
	TBM	0.7757	*	51
	FM	0.6055	*	24
C-I-TASSER	All	0.6073	1.01E-13	64
	TBM	0.7103	1.44E-07	49
	FM	0.4570	8.00E-10	15
I-TASSER	All	0.5349	1.08E-14	51
	TBM	0.6895	5.41E-08	45
	FM	0.3093	5.09E-11	6

Table S3. Comparison between the first models built by D-QUARK, C-QUARK and QUARK for the 91 CASP14 domain targets. The "Type" column indicates the category of each target. P-values were calculated between the TM-scores of the D-I-TASSER models and the model from the other two methods using one-sided Student's t-tests to test if DQUARK has significantly higher TM-scores. "#{TM-score>0.5}" represents the number of targets with TM-scores greater than 0.5.

Method	Type	TM-score	P-value	#{TM-score>0.5}
D-QUARK	All	0.7039	*	74
	TBM	0.7694	*	50
	FM	0.6084	*	24
C-QUARK	All	0.5881	1.17E-13	59
	TBM	0.7115	6.23E-08	48
	FM	0.4080	1.51E-09	11
QUARK	All	0.5289	3.95E-14	49
	TBM	0.6951	1.55E-07	47
	FM	0.2864	8.95E-08	2

Table S4. Comparison between the MSAs built by DeepMSA2 and DeepMSA for the 91 CASP14 domain targets. The "Type" column indicates the category of each target. The "SeqNum" column indicates the number of homologous sequences in the MSA, "*Neff*" represents the number of effective sequences. P-values were calculated between the "SeqNum" (or *Neff*) of the MSA built by DeepMSA2 and DeepMSA, using one-sided Student's t-tests to test if the DeepMSA2 alignments were significantly deeper.

Type	Method	SeqNum	P-value	*Neff*	P-value
All	DeepMSA2	7169	*	280	*
	DeepMSA	4379	6.54E-09	187	3.71E-09
TBM	DeepMSA2	10551	*	412	*
	DeepMSA	6754	4.09E-05	291	3.23E-05
FM	DeepMSA2	2233	*	88	*
	DeepMSA	913	2.37E-07	35	7.65E-07

Table S5. Comparison between the MAE_n of DeepPotential utilizing MSAs from DeepMSA2 and DeepMSA for the 91 CASP14 domain targets. The "Type" column indicates the category of each target. The "MAE_n" column represents the mean absolute distance error between the distances predicted by DeepPotential and distances calculated from the experimental structures (see **Text S1**). P-values were calculated between the MAE_n of DeepPotential using the MSAs built by DeepMSA2 and DeepMSA using one-sided Student's t-tests to test if DeepMSA2 resulted in significantly lower MAE_n.

Type	Method	MAE_n	P-value
All	DeepPotential with DeepMSA2	1.720	*
	DeepPotential with DeepMSA	2.383	5.36E-06
TBM	DeepPotential with DeepMSA2	1.459	*
	DeepPotential with DeepMSA	1.791	3.76E-02
FM	DeepPotential with DeepMSA2	2.102	*
	DeepPotential with DeepMSA	3.246	4.88E-06

Table S6. Comparison between the first models built by D-I-TASSER using MSAs from DeepMSA2 and DeepMSA for the 91 CASP14 domain targets. The "Type" column indicates the category of each target. P-values were calculated between the TM-scores of the D-I-TASSER models using MSAs from DeepMSA2 and DeepMSA by one-sided Student's t-tests to test if DeepMSA2 resulted in significantly higher TM-scores. "#{TM-score>0.5}" represents the number of targets with TM-scores greater than 0.5

Type	Method	TM-score	P-value	#{TM>0.5}
All	D-I-TASSER with DeepMSA2	0.7140	*	77
	D-I-TASSER with DeepMSA	0.6412	2.30E-08	67
TBM	D-I-TASSER with DeepMSA2	0.7884	*	53
	D-I-TASSER with DeepMSA	0.7488	1.79E-04	50
FM	D-I-TASSER with DeepMSA2	0.6055	*	24
	D-I-TASSER with DeepMSA	0.4842	3.71E-06	17

Table S7. Comparison between the first models built by D-QUARK using MSAs from DeepMSA2 and DeepMSA for the 91 CASP14 domain targets. The "Type" column indicates the category of each target. P-values were calculated between the TM-scores of the QUARK models using MSAs from DeepMSA2 and DeepMSA by one-sided Student's t-tests to test if DeepMSA2 resulted in significantly higher TM-scores. "#{TM-score>0.5}" represents the number of targets with TM-scores greater than 0.5

Type	Method	TM-score	P-value	#{TM>0.5}
All	D-QUARK with DeepMSA2	0.7114	*	76
	D-QUARK with DeepMSA	0.6212	6.83E-11	66
TBM	D-QUARK with DeepMSA2	0.7819	*	52
	D-QUARK with DeepMSA	0.7213	8.27E-07	49
FM	D-QUARK with DeepMSA2	0.6084	*	24
	D-QUARK with DeepMSA	0.4751	1.86E-06	17

Table S8. Domain partition information for the 17 CASP14 multi-domain targets. The "Target" column is the name of each target. The "CASP domain partition" columns include the number of domains and the domain boundaries given by the CASP14 assessors for each target. The "Zhang-Group domain partition" columns include the number of domains and the domain boundaries predicted by ThreaDom or FUpred for each target. The "NDO" is the normalized domain overlap score (NDO-score). The last four columns show the TM-scores of the first full-length models built by D-I-TASSER/D-QUARK and the average TM-score of the first domain-level models built by D-I-TASSER/D-QUARK for each target, respectively.

Target	CASP domain partition		Zhang-Group domain partition		NDO	TM-score of full-length model		Average TM-score of domain models	
	No.	Boundary	No.	Boundary		D-I-TASSER	D-QUARK	D-I-TASSER	D-QUARK
T1024	2	2-194; 203-406;	1	2-406;	0.509	0.8275	0.8303	0.8710	0.8739
T1030	2	1-154; 155-273;	1	1-273;	0.550	0.2332	0.2041	0.3332	0.2916
T1038	2	9-122; 123-198;	2	9-120; 121-198;	1.000	0.3005	0.3116	0.4718	0.4589
T1047s2	3	32-178; 20-31,179-249; 250-365;	3	20-30,176-247; 31-175; 248-365;	1.000	0.5239	0.5214	0.7734	0.7728
T1050	3	4-324; 325-640; 641-768;	3	4-320; 321-639; 640-768;	0.976	0.8795	0.8549	0.8679	0.8544
T1052	3	1-539; 540-588,669-832; 589-668;	3	1-540; 541-585,684-832; 586-683;	1.000	0.6617	0.6653	0.8156	0.8086
T1053	2	2-406; 407-577;	2	2-406; 397-577;	0.983	0.6188	0.7056	0.7344	0.7774
T1058	2	1-41,117-238,325-382; 42-116,239-324;	3	1-31,129-233,328-382; 32-128; 234-327;	0.650	0.7307	0.7133	0.8190	0.8247
T1061	3	1-170,442-735; 171-441; 736-838;	5	1-170; 160-455; 445-580; 570-735; 725-838;	0.671	0.2953	0.2937	0.5630	0.5515
T1070	4	4-79; 80-180; 181-256; 265-332;	4	4-90; 91-181; 182-262; 263-332;	0.942	0.3806	0.3843	0.7021	0.7060

T1091	4	359-497; 498-604; 605-710; 711-822;	4	359-498; 499-603; 604-710; 711-822;	1.000	0.471	0.4979	0.8227	0.8250
T1092	2	1-245; 246-426;	2	1-245; 246-426;	1.000	0.818	0.8033	0.7720	0.7665
T1093	3	3-143; 144-401,508-631; 402-507;	4	3-142; 143-265; 266-414,507-631; 415-506;	0.728	0.6262	0.6119	0.7229	0.7371
T1094	2	1-126,334-484; 127-333;	2	1-143,298-496; 144-297;	0.755	0.7429	0.7346	0.7793	0.7735
T1096	2	6-260; 294-464;	2	6-269; 270-464;	1.000	0.5181	0.5384	0.8337	0.7932
T1100	2	1-53,220-337; 54-219;	2	1-52,224-337; 53-223;	0.945	0.6461	0.6712	0.7204	0.7222
T1101	2	12-94; 95-318;	2	12-93; 94-318;	1.000	0.6311	0.6262	0.8559	0.8459
Average					0.865	0.5830	0.5860	0.7328	0.7284