

# Supplementary Materials

**Table S1** Proposed distance evaluation metrics.

Metrics	Definition	Applied case <sup>a</sup>	Category <sup>b</sup>	Range
<b>Contact precision (CP)</b>	$CP = \frac{1}{ S } \sum_{(i,j) \in S} I(D_{ij} \leq 8)$	P	C	[0,1]
<b>Absolute error (AE)</b>	$AE = \frac{1}{ S } \sum_{(i,j) \in S}  D_{ij} - d_{ij} $	P	R	[0,∞]
<b>Relative error (RE)</b>	$RE = \frac{1}{ S } \sum_{(i,j) \in S} \frac{ D_{ij} - d_{ij} }{D_{ij}}$	P	R	[0,∞]
<b>Pearson's correlation coefficient (PCC)</b>	$PCC = \frac{Cov(D_s, d_s)}{\sqrt{Var(D_s)Var(d_s)}}$	P	R	[-1,1]
<b>Distance precision (DP)</b>	$DP = \frac{1}{ S } \sum_{(i,j) \in S} P(d_{ij} \leq 20)I( D_{ij} - d_{ij}  \leq 2)$	P, N	R	[0,1]
<b>Fuzzy certainty (FC)</b>	$FC = \frac{1}{ S } \sum_{(i,j) \in S, D_{ij} \leq 20} P(D_{ij}) + \frac{1}{2}P(D_{ij} - 2)I(D_{ij} > 2) + \frac{1}{2}P(D_{ij} + 2)I(D_{ij} \leq 18)$	P, N	C	[0,1]
<b>Macro fuzzy precision (MFP) <sup>c</sup></b>	$fPRE_k = \begin{cases} \frac{1}{ S_k } \sum_{(i,j) \in S_k} I(l_{ij} = k) + \frac{1}{2}I( l_{ij} - k  = 1)I(D_{ij} \leq 20), & k < 10 \\ \frac{1}{ S_k } \sum_{(i,j) \in S_k} I(l_{ij} = k), & k = 10 \end{cases}$ $MFP = \frac{1}{n} \sum_{k=1}^n fPRE_k$	P, N, F	C	[0,1]
<b>Macro fuzzy recall (MFR) <sup>c</sup></b>	$fREC_k = \begin{cases} \frac{1}{N_k} \sum_{(i,j) \in S_k} I(l_{ij} = k) + \frac{1}{2}I( l_{ij} - k  = 1)I(D_{ij} \leq 20), & k < 10 \\ \frac{1}{N_k} \sum_{(i,j) \in S_k} I(l_{ij} = k), & k = 10 \end{cases}$ $MFR = \frac{1}{n} \sum_{k=1}^n fREC_k$	P, N, F	C	[0,1]
<b>Macro fuzzy F1 (MFF) <sup>c</sup></b>	$MFF = \frac{1}{n} \sum_{k=1}^n \frac{2fPRE_k fREC_k}{fPRE_k + fREC_k} I(fPRE_k + fREC_k > 0)$	P, N, F	C	[0,1]
<b>Distogram LDDT (DLDDT)</b>	$DLDDT = \frac{1}{4L} \sum_{t \in \{0.5, 1, 2, 4\}} \sum_{i=1}^L \frac{\sum_{j \in R_i} P(d_{ij} \leq 20)I( D_{ij} - d_{ij}  \leq t)}{ R_i } I( R_i  > 0)$	N	R	[0,1]
<b>Macro fuzzy certainty (MFC) <sup>c</sup></b>	$FC_k = \begin{cases} \frac{1}{ S_k } \sum_{(i,j) \in S_k} P_k(i, j) + \frac{1}{2}P_{k-1}(i, j)I(k > 1) + \frac{1}{2}P_{k+1}(i, j)I(k < 9), & k < 10 \\ \frac{1}{ S_k } \sum_{(i,j) \in S_k} P_k(i, j), & k = 10 \end{cases}$ $MFC = \frac{1}{n} \sum_{k=1}^n FC_k$	F	C	[0,1]

<sup>a</sup>: 'P', 'N', 'F' refer to the prediction-oriented, native-oriented, full-list case, respectively.

<sup>b</sup>: 'C' and 'R' refer to classification and regression, respectively.

<sup>c</sup>:  $n$  is the number of non-empty classes, with a maximum of 9 in the predicted- or native-oriented case and 10 in the full-list case.

**Table S2** Absolute values and summed Z-scores of the Pearson's correlation coefficient between the distance evaluation metrics and the TM-score (the native-oriented case). The highest value in each column is highlighted in bold type. The metrics are ranked based on the summed Z-score.

Metrics	D680	CASP13-31	CAMEO131	Summed Z-score
<b>DLDDT</b>	0.674	<b>0.695</b>	0.666	<b>0.963</b>
<b>DP</b>	0.676	0.666	<b>0.678</b>	0.936
<b>MF</b>	<b>0.68</b>	0.647	0.655	0.717
<b>FC</b>	0.653	0.645	0.646	-0.365
<b>MFR</b>	0.648	0.613	0.637	-0.858
<b>MFP</b>	0.648	0.612	0.586	-1.393

**Table S3** Absolute values and summed Z-scores of the Pearson's correlation coefficient between the distance evaluation metrics and the TM-score (the full-list case). The highest value in each column is highlighted in bold type. The metrics are ranked based on the summed Z-score.

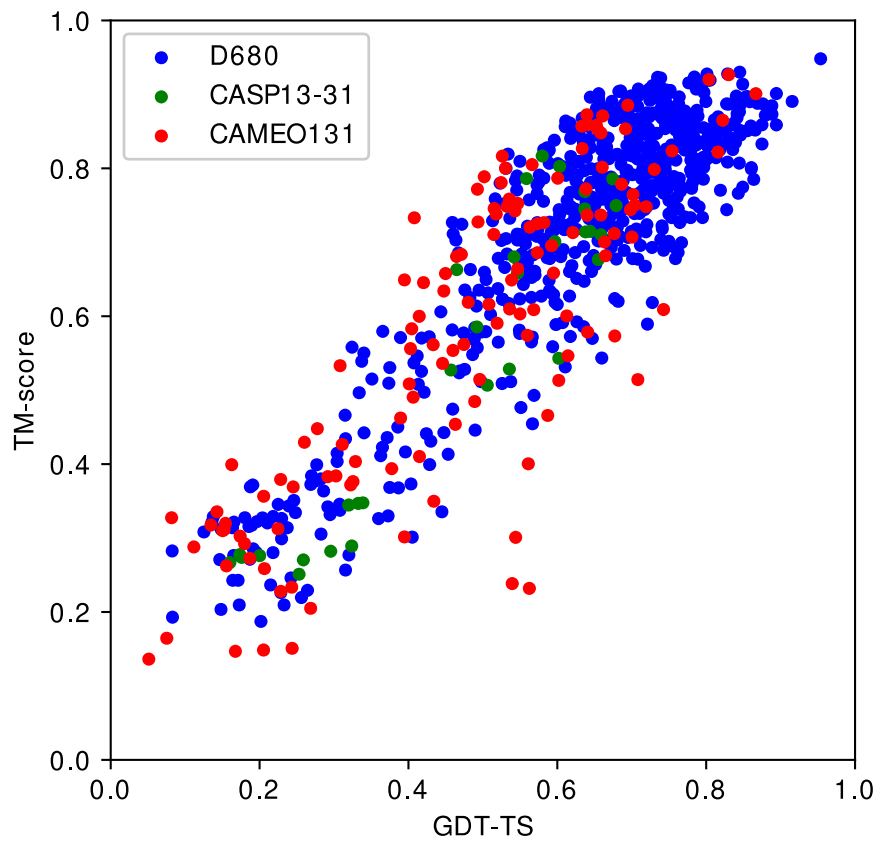
Metrics	D680	CASP13-31	CAMEO131	Summed Z-score
<b>MF</b>	<b>0.712</b>	<b>0.705</b>	<b>0.695</b>	<b>1.266</b>
<b>MFP</b>	0.694	0.701	0.649	-0.03
<b>MFC</b>	0.69	0.607	0.678	-0.107
<b>MFR</b>	0.663	0.634	0.655	-1.129

**Table S4** Absolute values of the Pearson's correlation coefficient between *P20* and *mP20* with distance evaluation metrics in the prediction-oriented case on D680. The highest value in each column is highlighted in bold type.

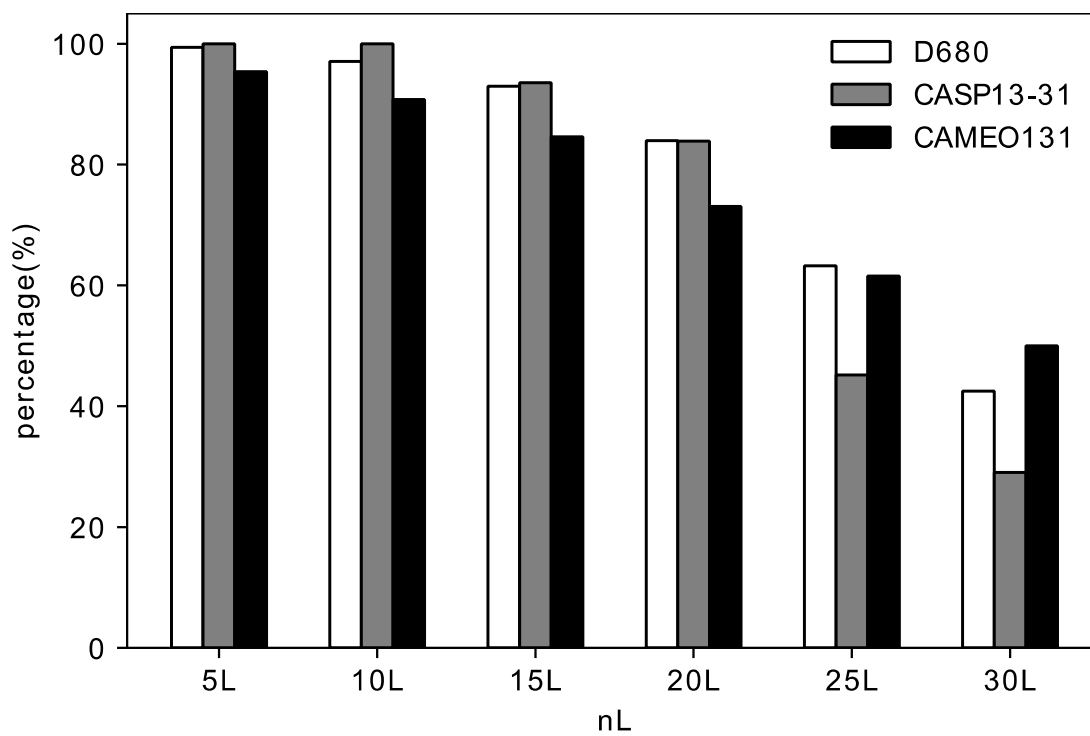
Metrics	<i>P20</i>	<i>mP20</i>
<b>FC</b>	<b>0.948</b>	<b>0.92</b>
<b>DP</b>	0.924	0.9
<b>RE</b>	0.827	0.822
<b>CP</b>	0.797	0.785
<b>MFP</b>	0.754	0.8
<b>PCC</b>	0.716	0.719
<b>AE</b>	0.681	0.668
<b>MF</b>	0.497	0.497
<b>MFR</b>	0.306	0.315

**Table S5** Rankings of distance prediction on 37 CASP14 FM domains. The group information can be found in CASP14 webpage (<https://predictioncenter.org/casp14/docs.cgi?view=groupsbyname>).

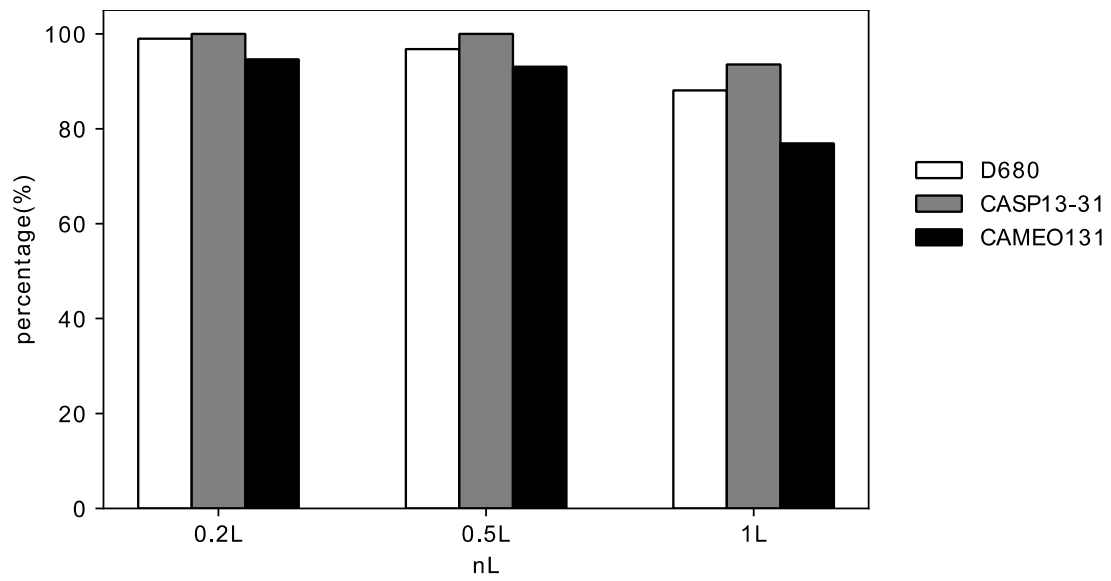
<b>Rank</b>	<b>Official</b>	<b>DP (prediction-oriented)</b>	<b>DLDDT (native-oriented)</b>	<b>MFF (full-list)</b>
1	488	368	304	368
2	009	024	024	488
3	368	010	010	024
4	024	488	368	010
5	010	009	488	009
6	304	304	192	183
7	192	192	009	351
8	351	351	351	238
9	183	183	140	125
10	140	140	183	304
11	367	367	409	192
12	377	409	453	409
13	326	377	367	140
14	222	222	222	367
15	409	326	377	326
16	238	238	326	222
17	319	198	238	377
18	198	319	275	319
19	125	275	319	198
20	453	252	198	420
21	252	420	252	275
22	420	125	420	252
23	275	187	071	187
24	061	307	125	061
25	187	453	364	487
26	307	392	307	453
27	392	071	392	307
28	364	364	187	392
29	071	061	061	071
30	487	487	487	364
31	387	151	151	151
32	254	254	254	157
33	151	388	323	388
34	388	157	157	254
35	425	323	361	323
36	323	425	144	425
37	144	144	388	361
38	157	361	425	144
39	361	387	387	387



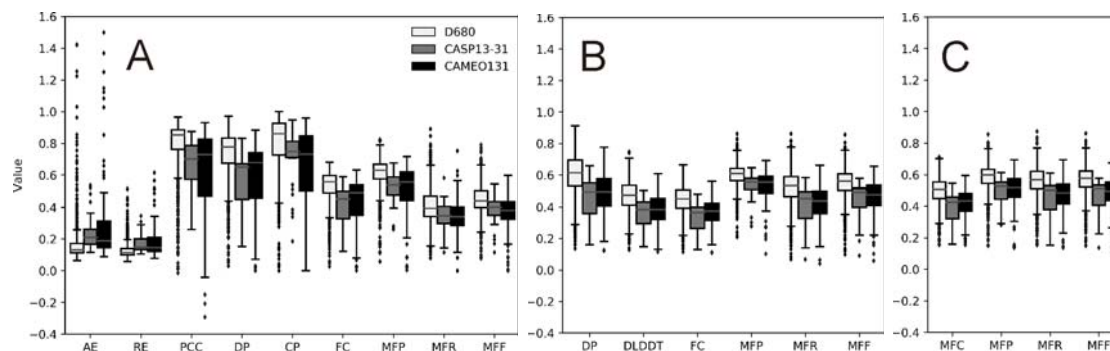
**Figure S1** Scatter plot of GDT-TS and TM-score of models built by trRosetta on D680 (blue dots), CASP13-31 (green dots), and CAMEO131 (red dots).



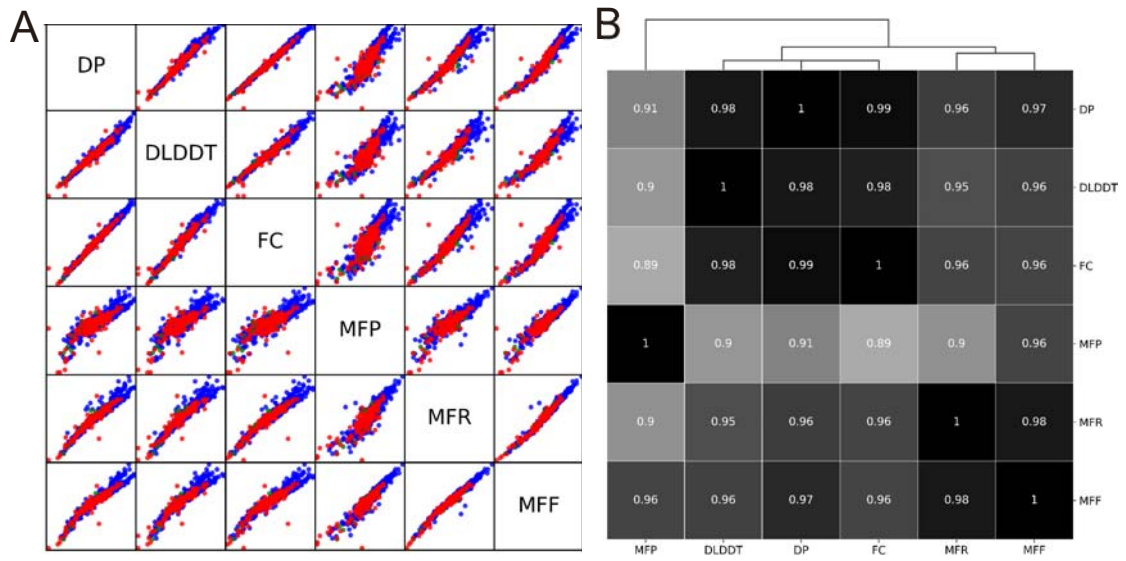
**Figure S2** Percentage of targets satisfying  $N \geq nL$ .  $N$  is the number of residue pairs with native distance  $\leq 20\text{\AA}$  and sequence separation  $\geq 12$ .  $L$  is the length of a target.



**Figure S3** The percentages of proteins satisfying  $N_c \geq nL$  ( $n=0.2, 0.5, 1$ ), where  $N_c$  denotes the number of long and medium contacts, and  $L$  is the length of a protein.

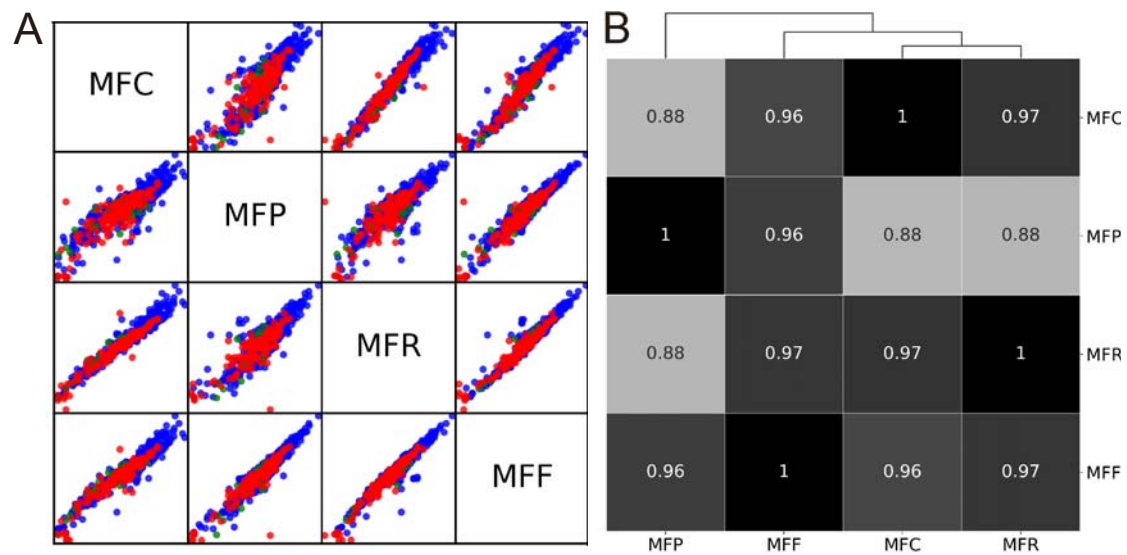


**Figure S4.** Boxplots of the proposed metrics on three datasets. (A-C) are for the prediction-oriented, the native-oriented and the full-list cases, respectively. The interquartile range (i.e., the box length) represents the range of the middle 50% values. The horizontal line within each box represents the median. Whiskers outside the box indicate the variability outside the upper and the lower quartiles. The black dots are the outliers.

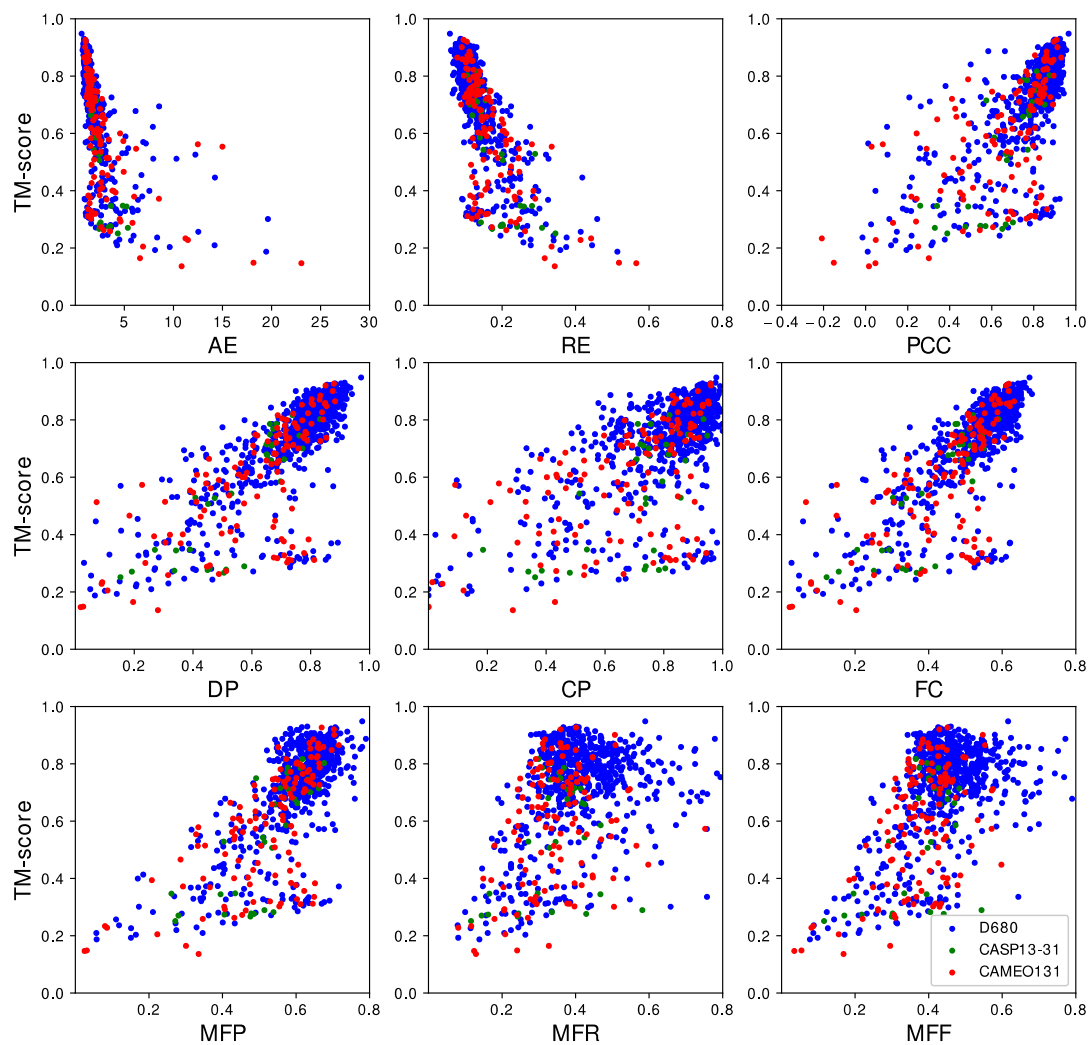


**Figure S5** Scatter matrix and Spearman's rank correlation matrix (absolute value) for distance evaluation metrics in the native-oriented case.

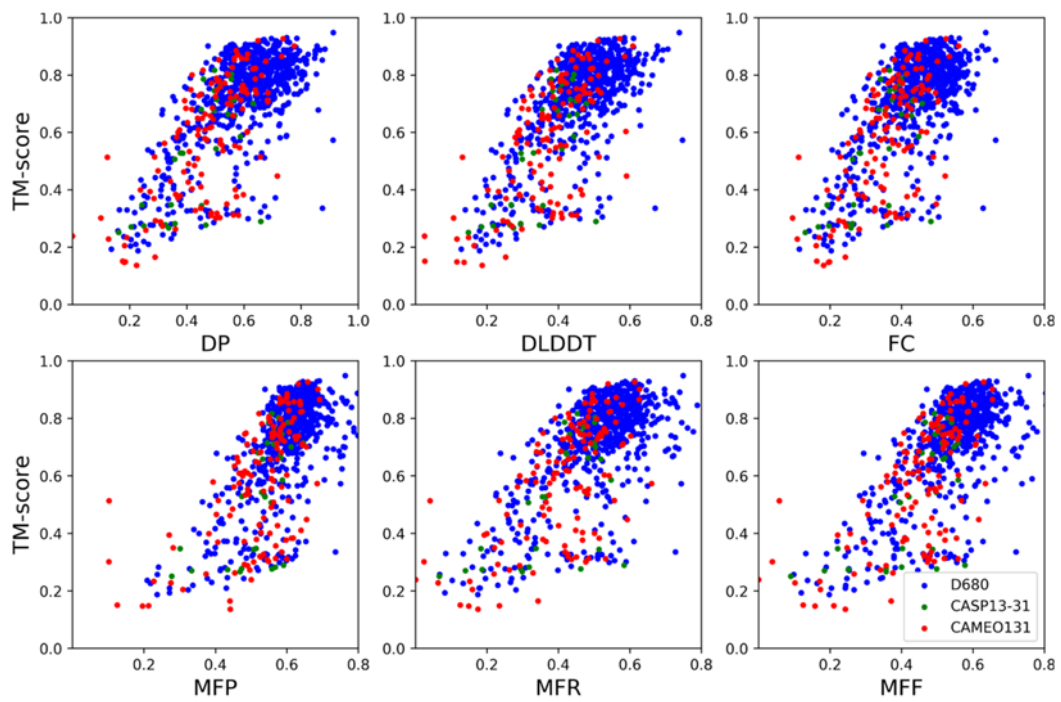




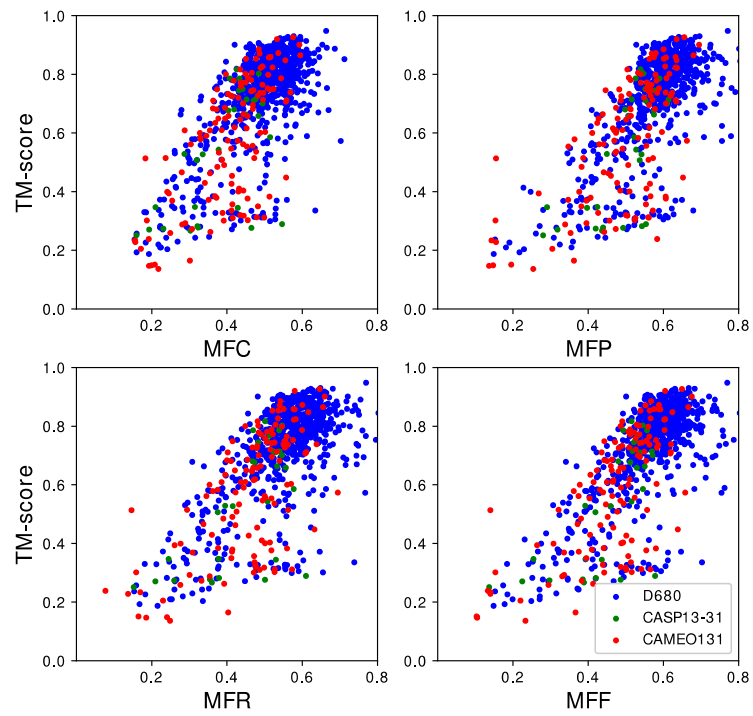
**Figure S6** Scatter matrix and Spearman's rank correlation matrix (absolute value) for distance evaluation metrics in the full-list case.



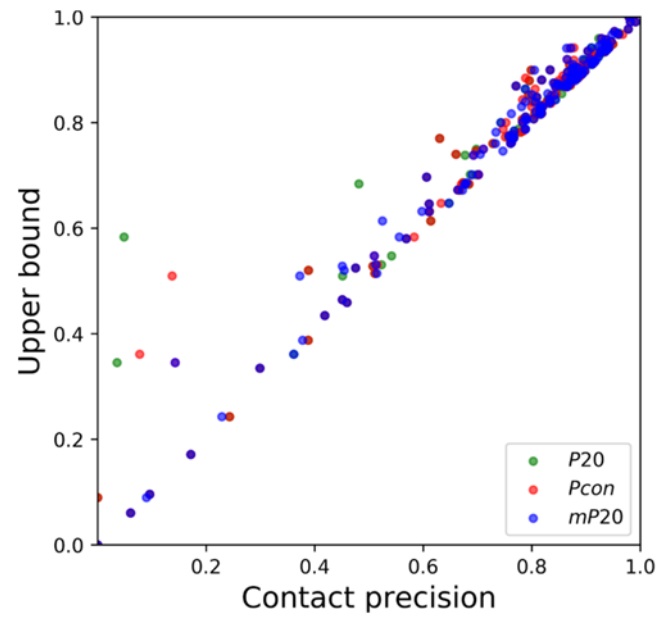
**Figure S7** Scatter plots of distance evaluation metrics in the prediction-oriented case and TM-score on D680 (blue dots), CASP13-31 (green dots) and CAMEO131 (red dots).



**Figure S8** Scatter plots of distance evaluation metrics in the native-oriented case and TM-score on D680 (blue dots), CASP13-31 (green dots) and CAMEO131 (red dots).



**Figure S9** Scatter plots of distance evaluation metrics in the full-list case and TM-score on D680 (blue dots), CASP13-31 (green dots) and CAMEO131 (red dots).



**Figure S10** Scatter plots of the upper bound (maximum contact precision out of all MSAs, y-axis) versus the actual contact precision (x-axis) by *Pcon* (red dots), *P20* (green dots) and *mP20* (blue dots) on 161 CAMEO targets.