Supplementary Materials for Protein Structure Prediction in CASP13 using AWSEM-Suite

Shikai Jin[1,2,*], Mingchen Chen[1,*], Xun Chen[1,3,*], Carlos Bueno[1], Wei Lu[1,4], Nicholas P. Schafer[1,3], Xingcheng Lin[5], José N. Onuchic[1,2,3,4], and Peter G. Wolynes[1,2,3,4,†]

[1]Center for Theoretical Biological Physics, Rice University, Houston, TX, USA
[2]Department of Biosciences, Rice University, Houston, TX, USA
[3]Department of Chemistry, Rice University, Houston, TX, USA
[4]Department of Physics, Rice University, Houston, TX, USA
[5]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA
[†]Please address correspondence to pwolynes@rice.edu.
[*]These authors contributed equally to this work.

May 4, 2020

Contents

S1 Qo value under different type of contacts for AWSEM-Suite and Baker-RosettaServer.

Table S1: Summary of Qo value for different type of contacts in AWSEM-Suite and Baker-RosettaServer predicted structures.

Domain name	Short contact AWSEM	Short contact Rosetta	Long contact AWSEM	Long contact Rosetta	Water-mediated contact AWSEM	Water-mediated contact Rosetta	Water-mediated contact proportion
T0950-D1	0.687	0.684	0.508	0.596	0.654	0.638	0.052
T0951-D1	0.915	0.966	0.899	0.959	0.869	0.938	0.048
T0953s1-D1	0.501	0.706	0.438	0.720	0.620	0.613	0.225
T0953s2-D1	0.374	0.461	0.320	0.438	0.592	0.645	0.08
T0953s2-D2	0.716	0.414	0.734	0.436	0.756	0.458	0.05
T0953s2-D3	0.343	0.168	0.372	0.270	0.645	0.643	0.013
T0954-D1	0.614	0.770	0.623	0.773	0.528	0.662	0.071
T0955-D1	0.630	0.000	0.530	0.000	0.579	0.000	0.212
T0957s1-D1	0.566	0.666	0.507	0.535	0.489	0.591	0.095
T0957s1-D2	0.741	0.839	0.603	0.649	0.811	0.790	0.082
T0957s2-D1	0.743	0.730	0.627	0.635	0.659	0.565	0.068
T0958-D1	0.827	0.849	0.755	0.751	0.830	0.748	0.030
T0960-D2	0.402	0.555	0.392	0.576	0.519	0.722	0.142
T0960-D3	0.565	0.767	0.538	0.766	0.598	0.731	0.213
T0960-D5	0.641	0.712	0.633	0.742	0.654	0.708	0.073
T0963-D2	0.408	0.389	0.396	0.385	0.588	0.552	0.168

T0963-D3	0.632	0.566	0.617	0.557	0.635	0.616	0.122
T0963-D5	0.674	0.656	0.668	0.674	0.676	0.648	0.082
T0965-D1	0.809	0.796	0.782	0.785	0.717	0.702	0.049
T0966-D1	0.752	0.825	0.732	0.797	0.731	0.785	0.051
T0967-D1	0.864	0.955	0.852	0.952	0.780	0.918	0.09
T0968s1-D1	0.690	0.843	0.609	0.790	0.570	0.775	0.099
T0968s2-D1	0.558	0.805	0.635	0.823	0.607	0.806	0.161
T0969-D1	0.532	0.482	0.451	0.457	0.480	0.467	0.108
T0970-D1	0.745	0.491	0.788	0.525	0.670	0.504	0.112
T0971-D1	0.943	0.960	0.914	0.968	0.965	0.959	0.04
T0976-D1	0.812	0.849	0.805	0.835	0.776	0.860	0.068
T0976-D2	0.862	0.832	0.836	0.833	0.832	0.781	0.084
T0980s1-D1	0.496	0.477	0.446	0.424	0.552	0.494	0.121
T0984-D1	0.708	0.822	0.650	0.790	0.611	0.595	0.066
T0984-D2	0.729	0.893	0.607	0.857	0.709	0.756	0.114
T0986s1-D1	0.689	0.804	0.653	0.766	0.701	0.693	0.075
T0986s2-D1	0.526	0.426	0.486	0.376	0.472	0.497	0.065
T0990-D1	0.805	0.634	0.695	0.547	0.832	0.576	0.11
T0990-D2	0.636	0.561	0.542	0.418	0.564	0.465	0.105
T0990-D3	0.507	0.445	0.432	0.391	0.594	0.410	0.087
T1003-D1	0.856	0.944	0.829	0.924	0.717	0.803	0.057
T1005-D1	0.613	0.652	0.585	0.638	0.583	0.557	0.063
T1006-D1	0.906	0.973	0.910	0.952	0.894	0.924	0.125
T1008-D1	0.632	0.821	0.565	0.762	0.694	0.764	0.104
T1011-D1	0.658	0.831	0.602	0.788	0.609	0.595	0.043
T1016-D1	0.857	0.907	0.835	0.897	0.768	0.822	0.066
T1018-D1	0.788	0.923	0.778	0.914	0.833	0.921	0.023

T1021s1-D1	0.577	0.870	0.587	0.890	0.628	0.780	0.222
T1021s2-D1	0.633	0.813	0.596	0.784	0.660	0.727	0.085
T1021s3-D1	0.553	0.620	0.524	0.581	0.545	0.585	0.129
T1021s3-D2	0.420	0.366	0.348	0.389	0.420	0.386	0.105
T1022s1-D1	0.378	0.580	0.418	0.617	0.529	0.525	0.173
T1022s1-D2	0.761	0.877	0.732	0.857	0.762	0.752	0.19
T1022s2-D1	0.550	0.804	0.547	0.790	0.606	0.717	0.138

S2 The quality of the template and coevolutionary contact pairs for 8 selected structures.

Table S2: Summary of the quality of the template and coevolutionary contact pairs for 8 selected structures.

Domain name	Coevolutionary method	Number of predicted contact pair	True positive rate of contact pair	Template name	Template sequence identity to the domain
T0970-D1	RaptorX-Contact	115	52.1%	1JG5_C	11%
T0976-D1	RaptorX-Contact	54	57.4%	1YT8_A	17%
T0976-D2	RaptorX-Contact	67	59.7%	1YT8_A	16%
T0984-D1	DeepContact	9121	2.5%	6C9A_A	24%

T1005-D1	RaptorX-Contact	524	42.3%	3W4R_A	14%
T1011-D1	DeepContact	3818	3.9%	3ODU_B	47%
T1018-D1	DeepContact	8892	4.7%	3IAR_A	31%
T1021s2-D1	DeepContact	10904	4.2%	3J9Q_A	19%

S3 Evaluation of Qo value over different distance cutoff for 4 groups.

S4 The contact map comparison of T0958-D1.

S5 Summary of the quality of energy-based blind selection from prediction trajectory.

S6 Free energy profiles of T0958 using AWSEM force field under 200K and 400K.

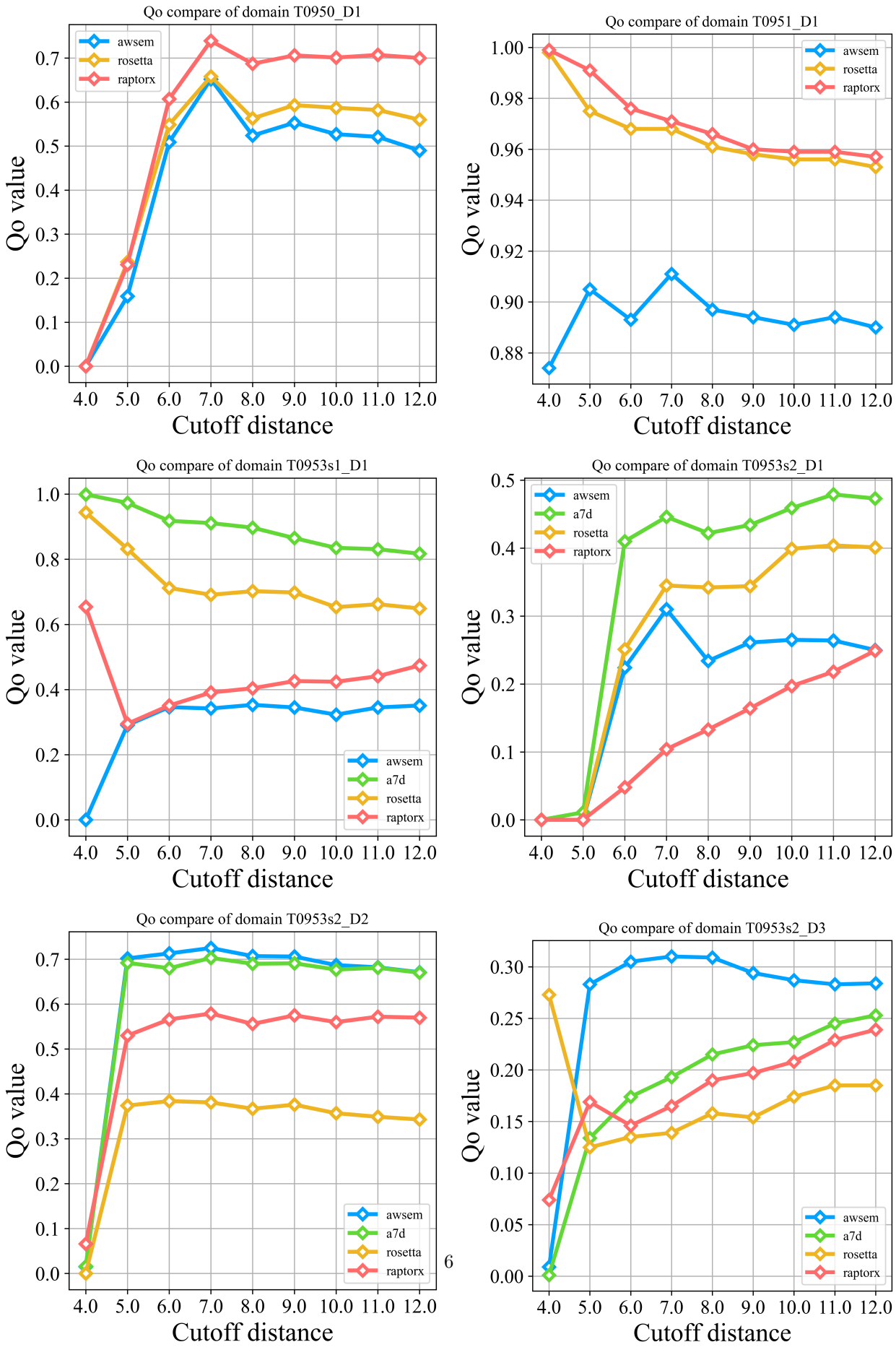
S7 Values of energy terms in different AWSEM as a function of Q for the protein T0958.

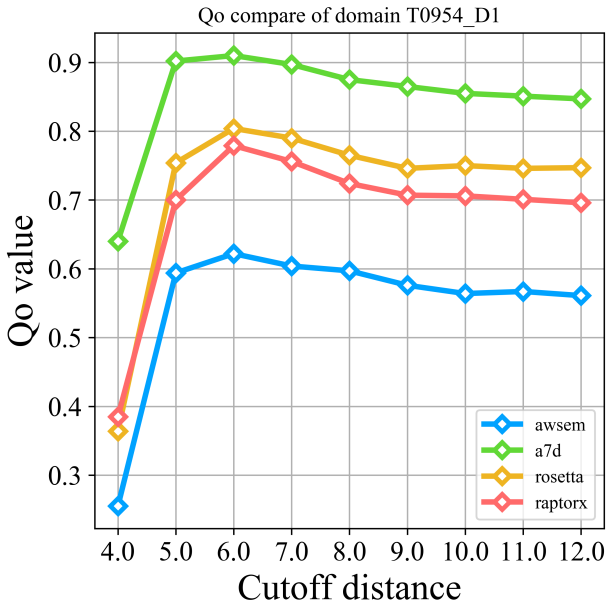
S8 Correct contact versus length and sequence identity.

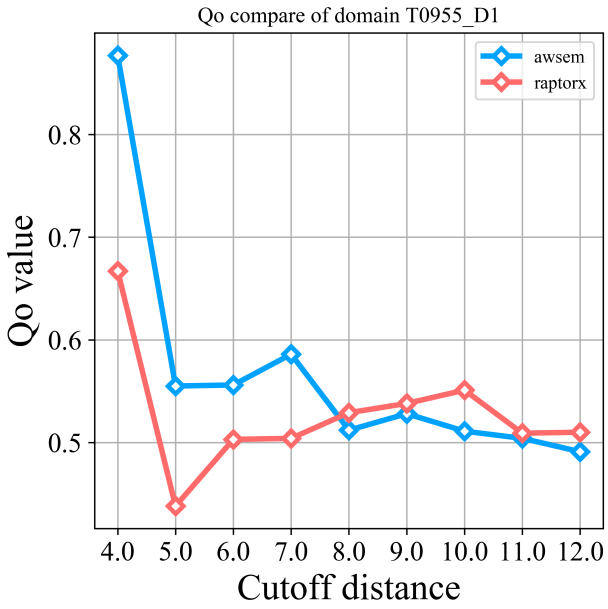
S9 Ratio of correct contacts in different parts divided by domain length versus the Qw value of each domain.

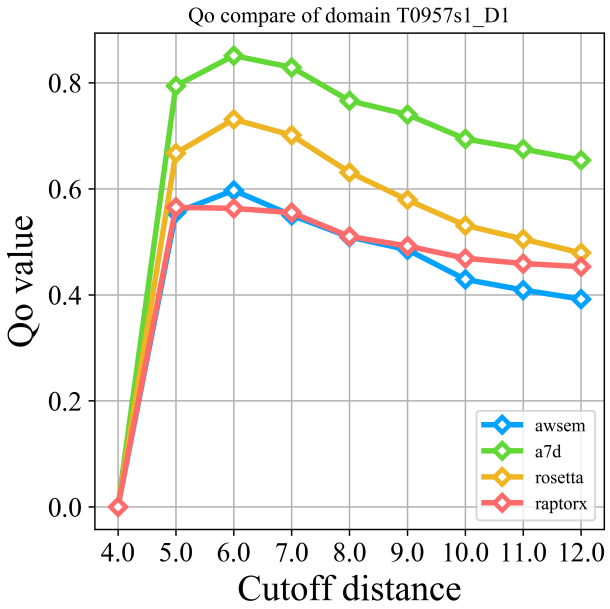
S10 Qw value versus the secondary structure percent.

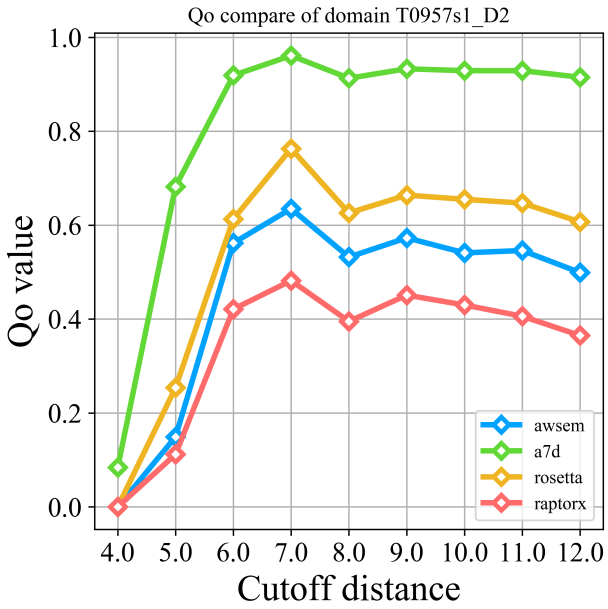
Figure S1: Evaluation of Qo value over different distance cutoffs for 4 groups on 50 domains.

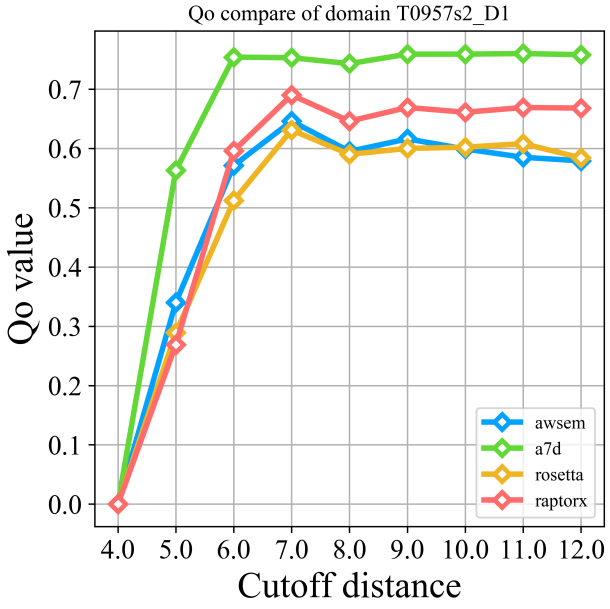


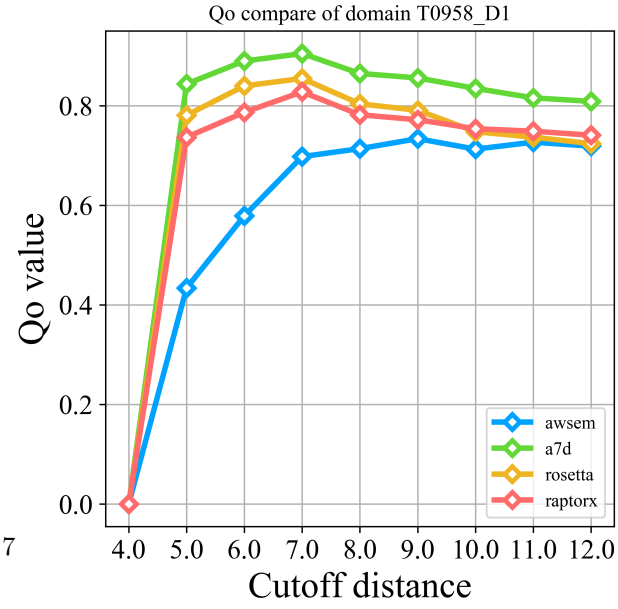

Qo compare of domain T0954_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

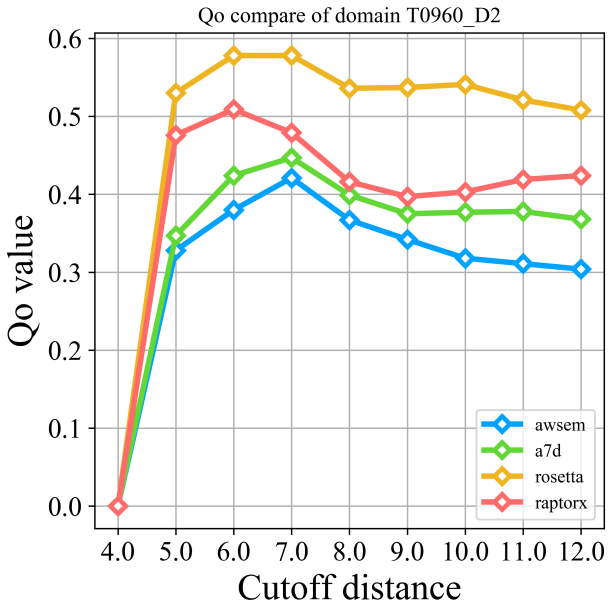

Qo compare of domain T0955_D1
Qo value
Cutoff distance
awsem
raptorx

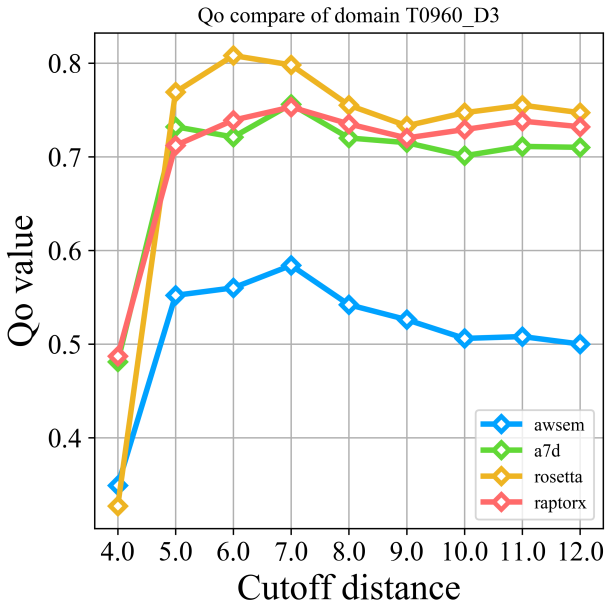

Qo compare of domain T0957s1_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

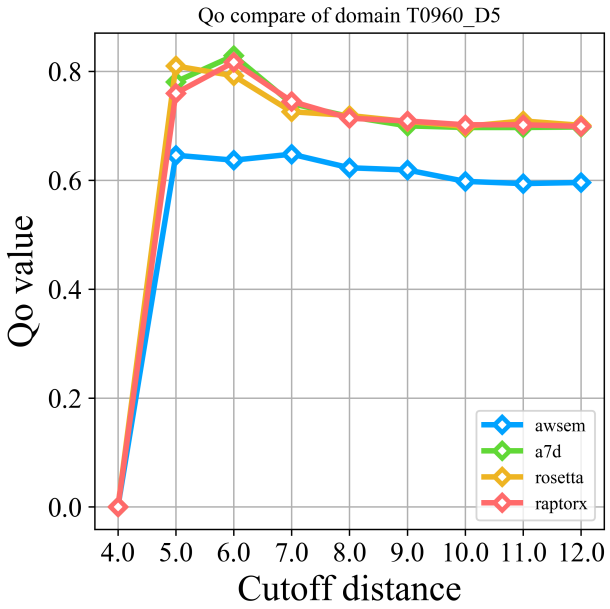

Qo compare of domain T0957s1_D2
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

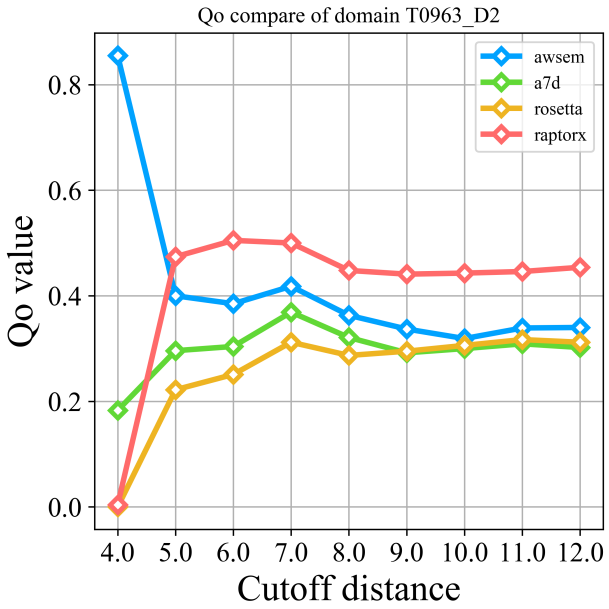

Qo compare of domain T0957s2_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

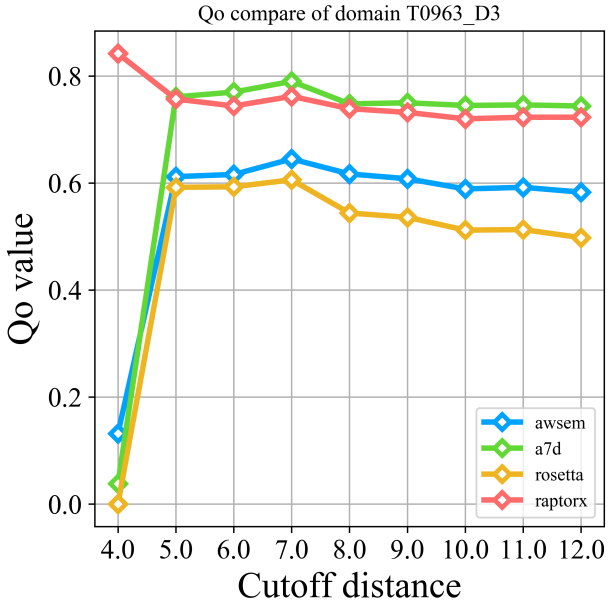

Qo compare of domain T0958_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

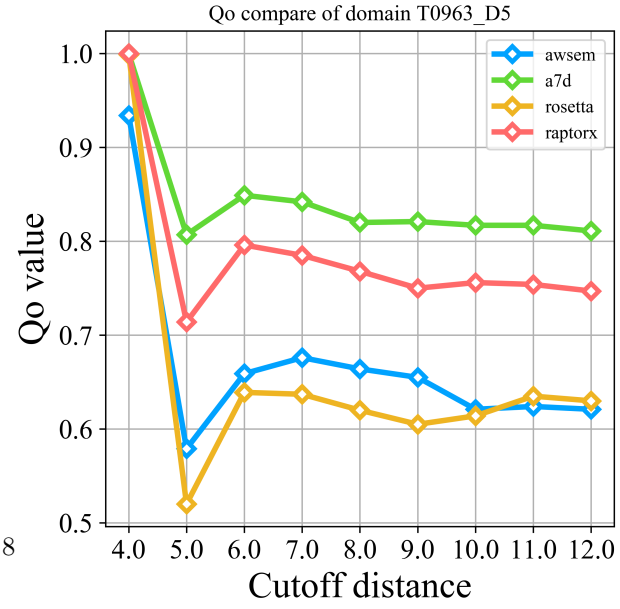

Qo compare of domain T0960_D2
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

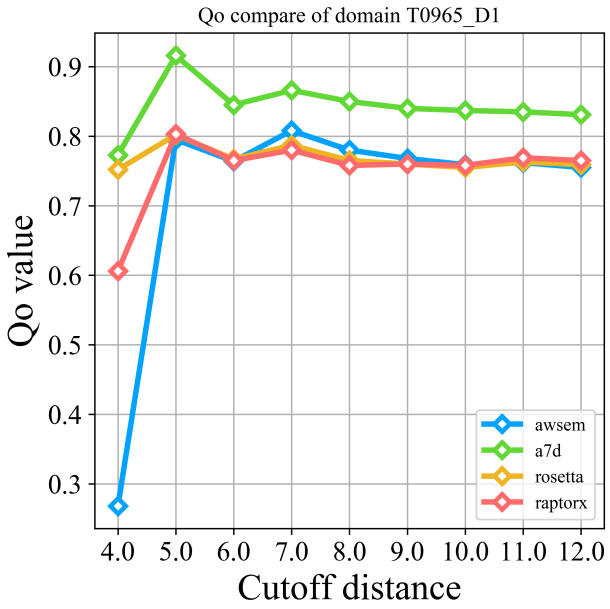

Qo compare of domain T0960_D3
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

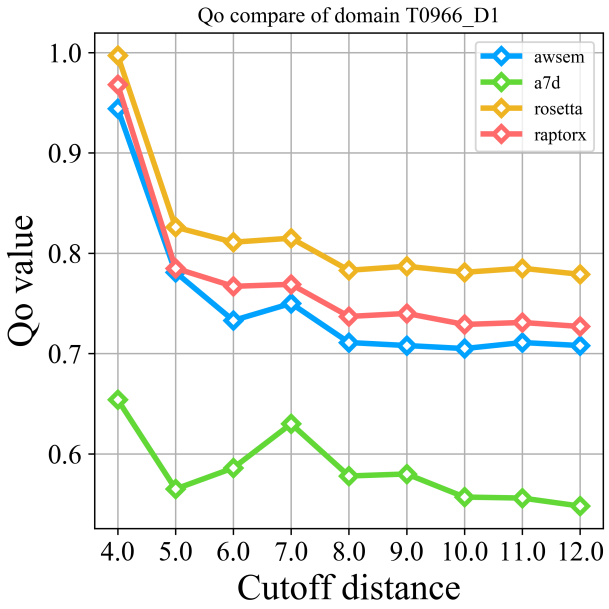

Qo compare of domain T0960_D5
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

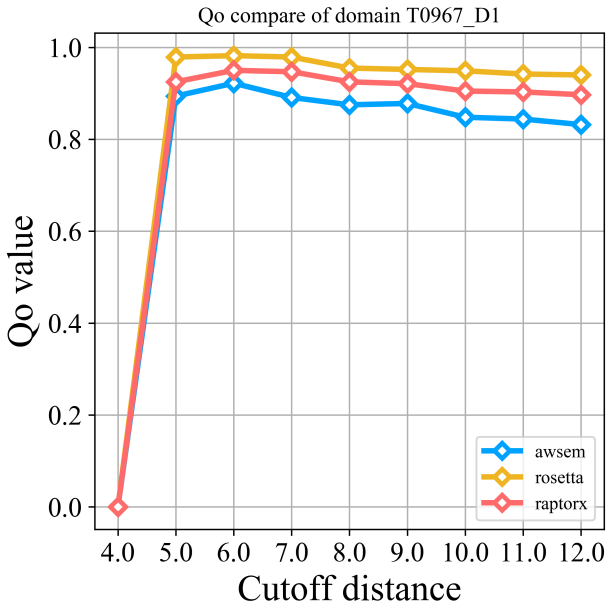

Qo compare of domain T0963_D2
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

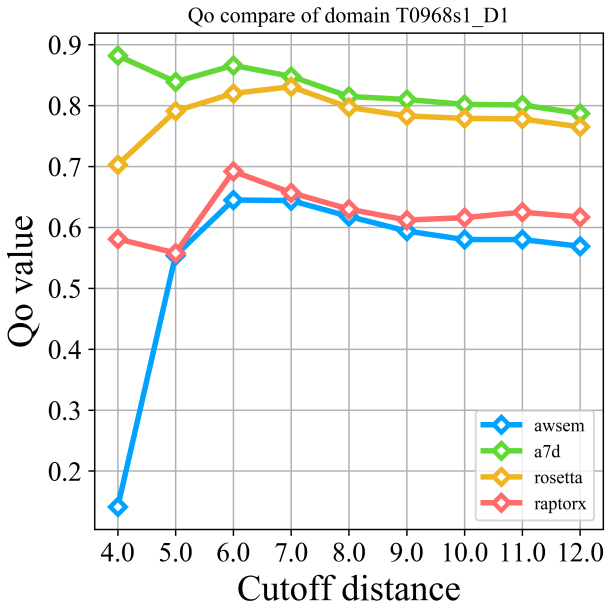

Qo compare of domain T0963_D3
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

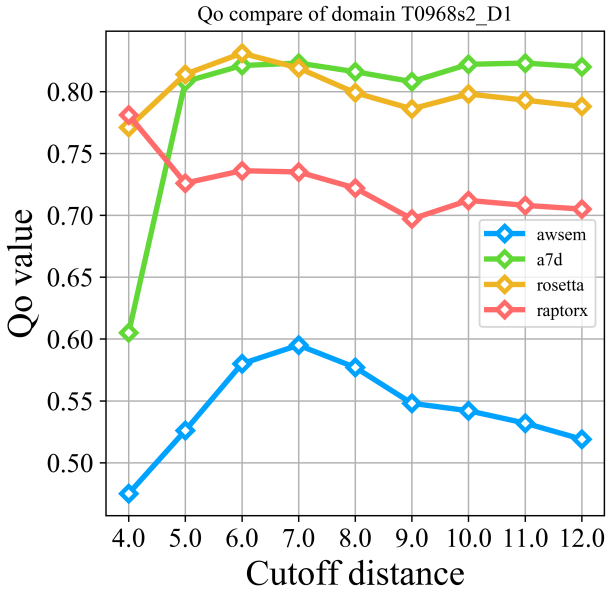

Qo compare of domain T0963_D5
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

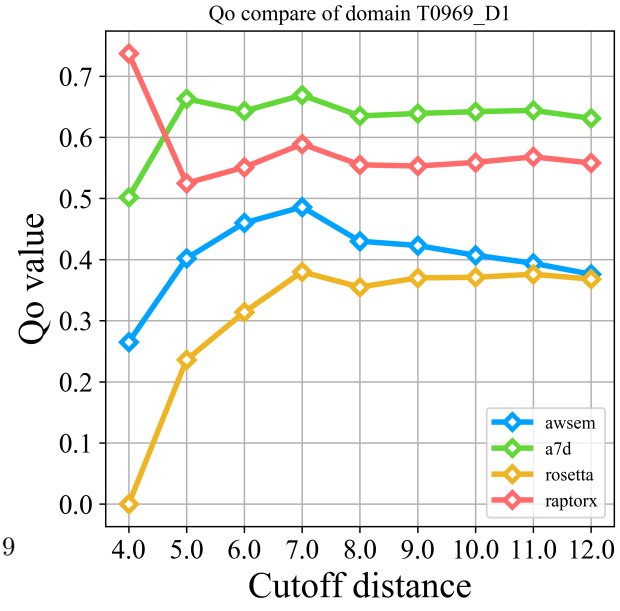

Qo compare of domain T0965_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

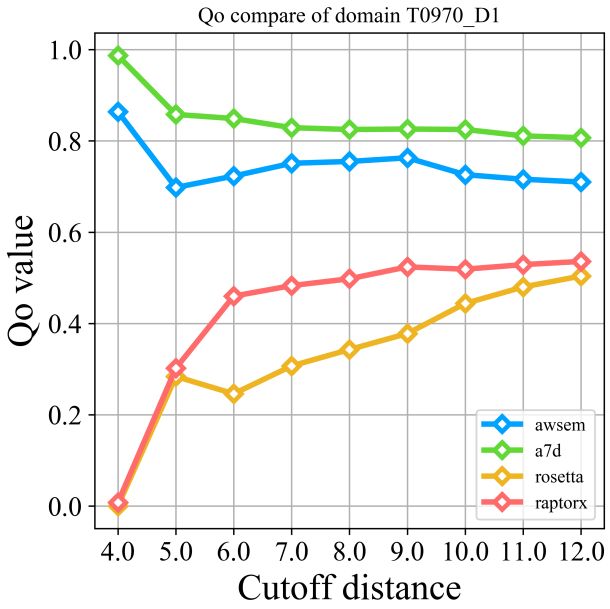

Qo compare of domain T0966_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

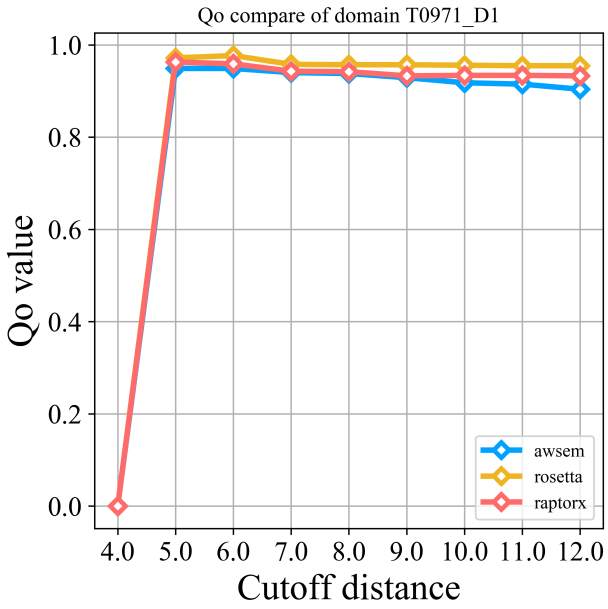

Qo compare of domain T0967_D1
Qo value
Cutoff distance
awsem
rosetta
raptorx

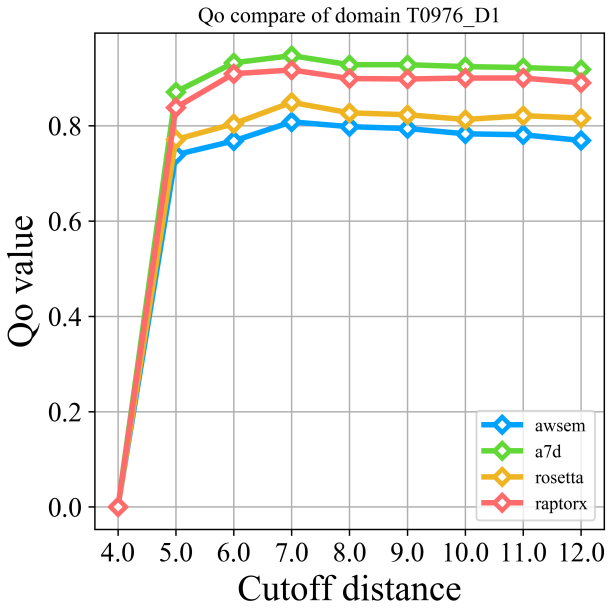

Qo compare of domain T0968s1_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

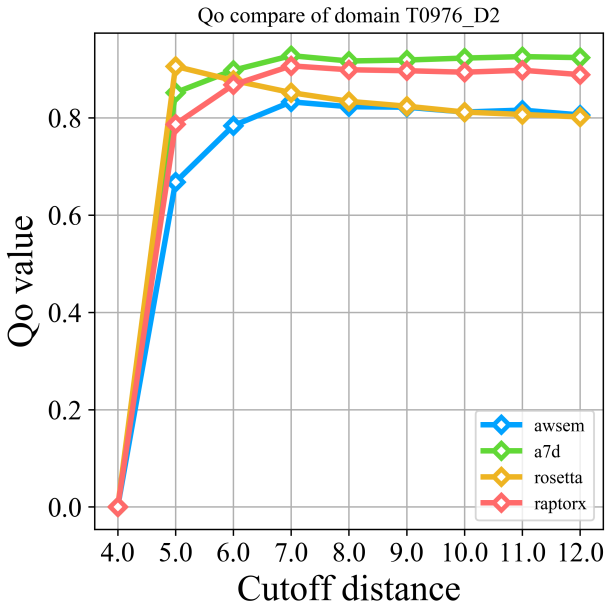

Qo compare of domain T0968s2_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

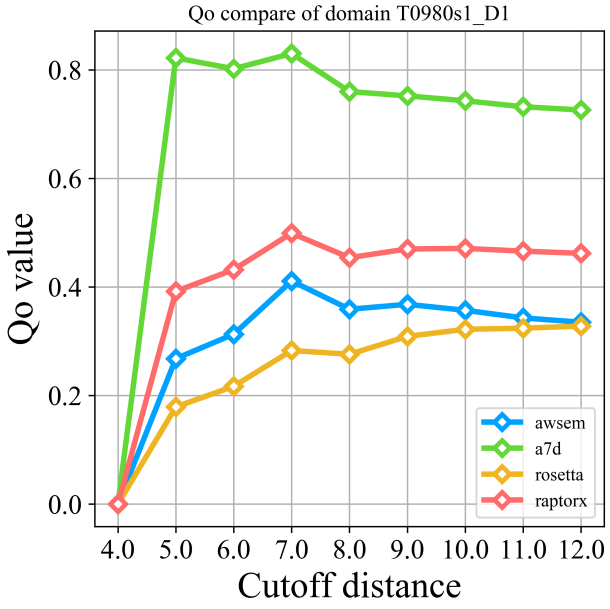

Qo compare of domain T0969_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

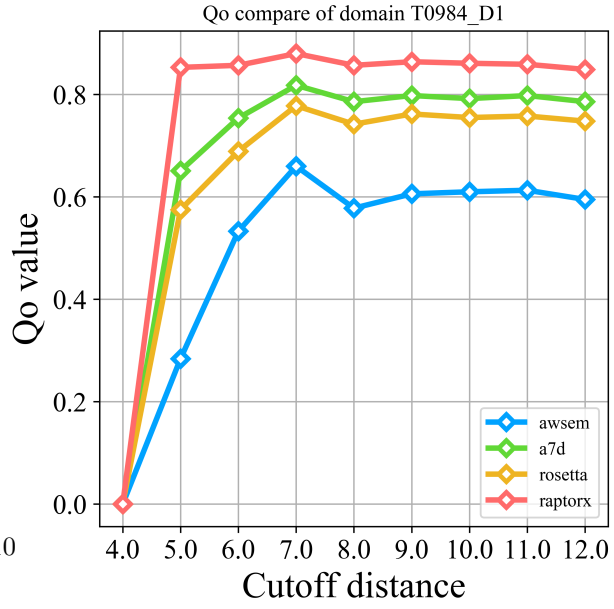

Qo compare of domain T0970_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

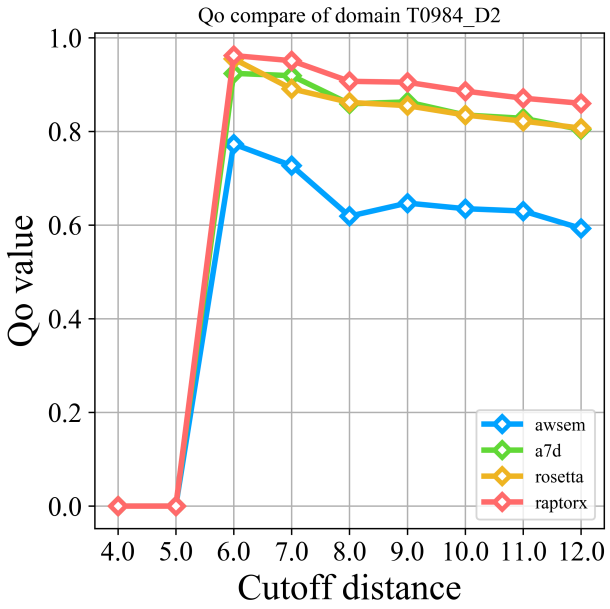

Qo compare of domain T0971_D1
Qo value
Cutoff distance
awsem
rosetta
raptorx

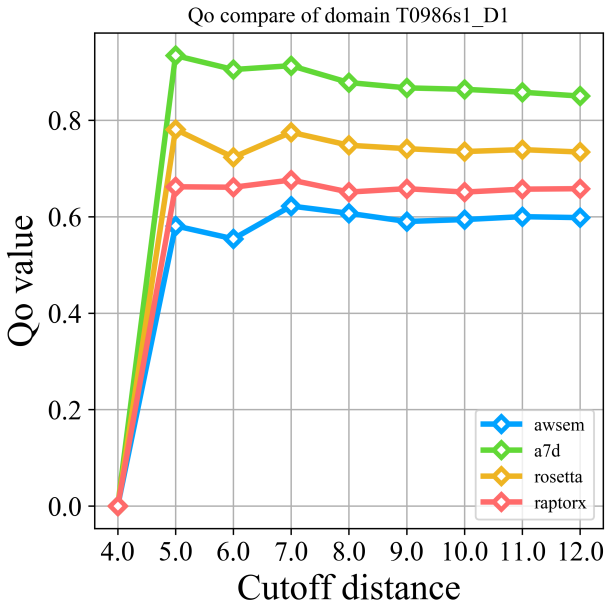

Qo compare of domain T0976_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

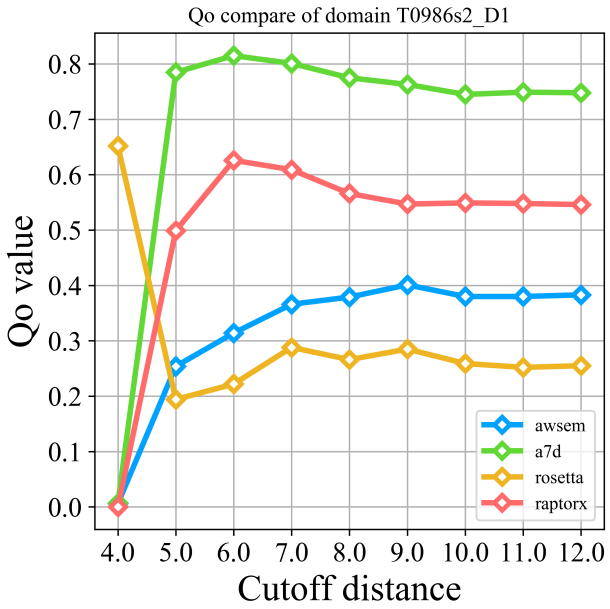

Qo compare of domain T0976_D2
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

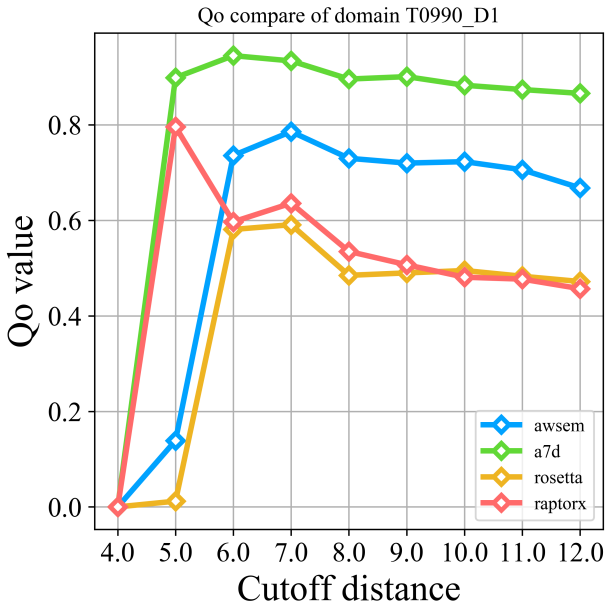

Qo compare of domain T0980s1_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

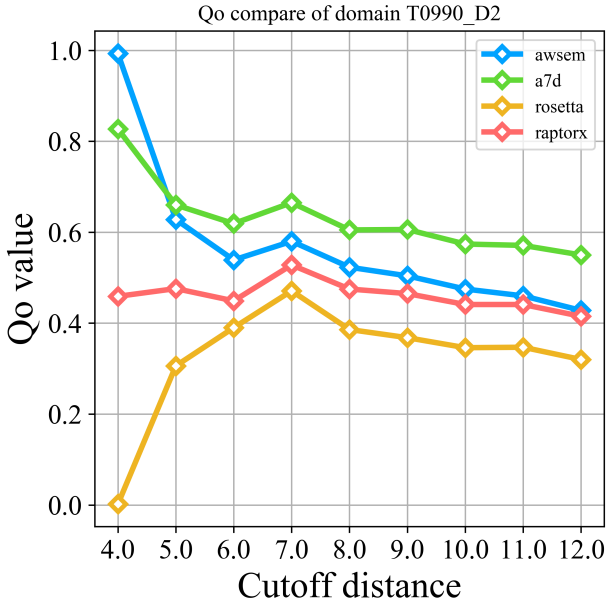

Qo compare of domain T0984_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

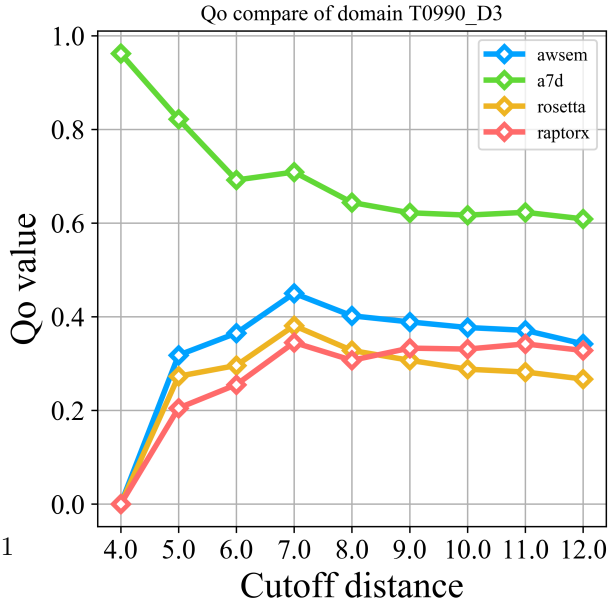

Qo compare of domain T0984_D2
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

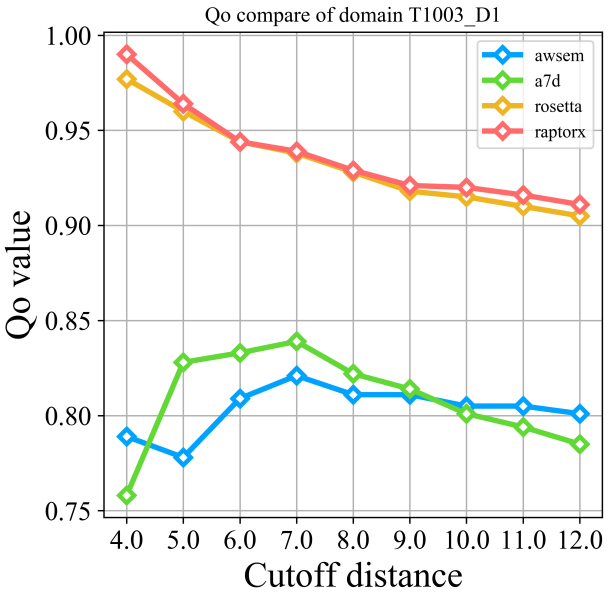

Qo compare of domain T0986s1_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

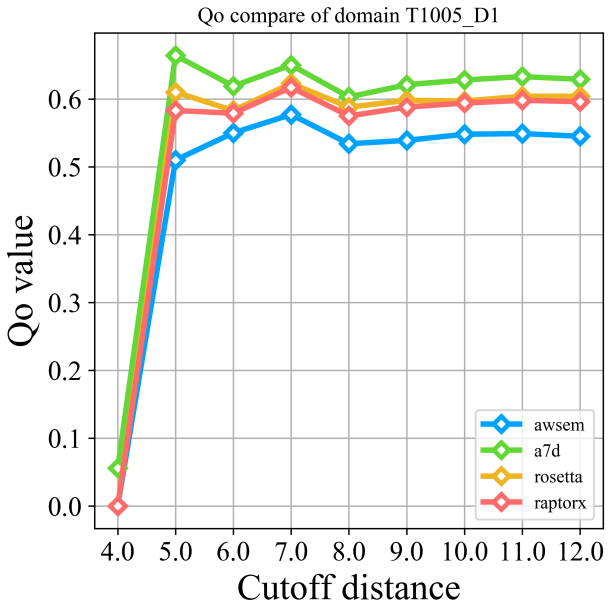

Qo compare of domain T0986s2_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

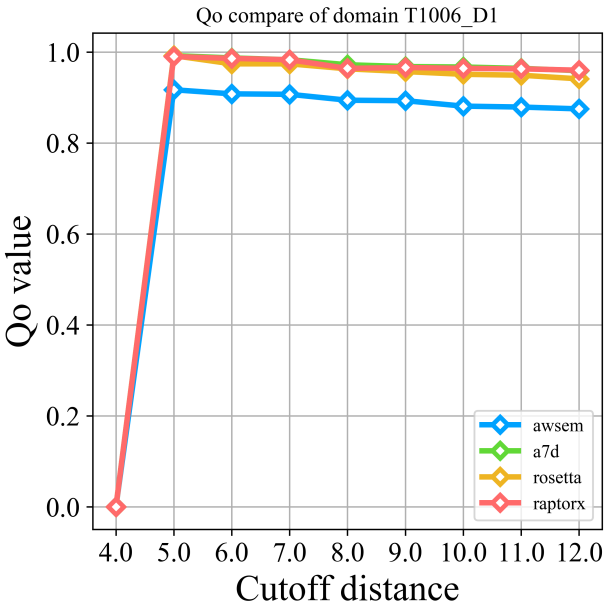

Qo compare of domain T0990_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

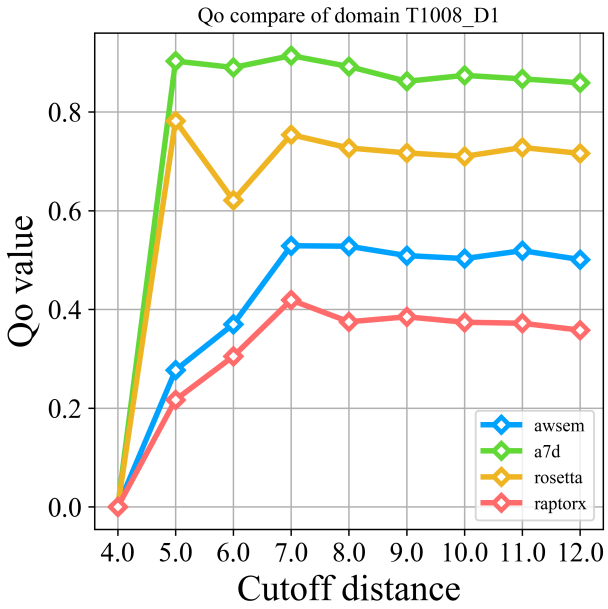

Qo compare of domain T0990_D2
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

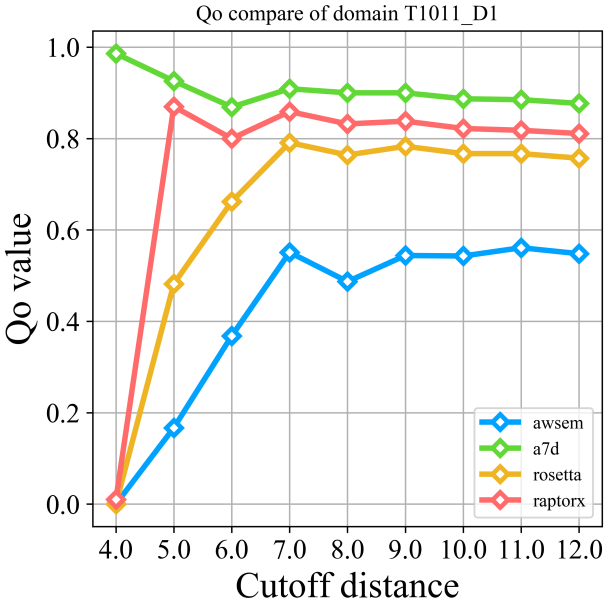

Qo compare of domain T0990_D3
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

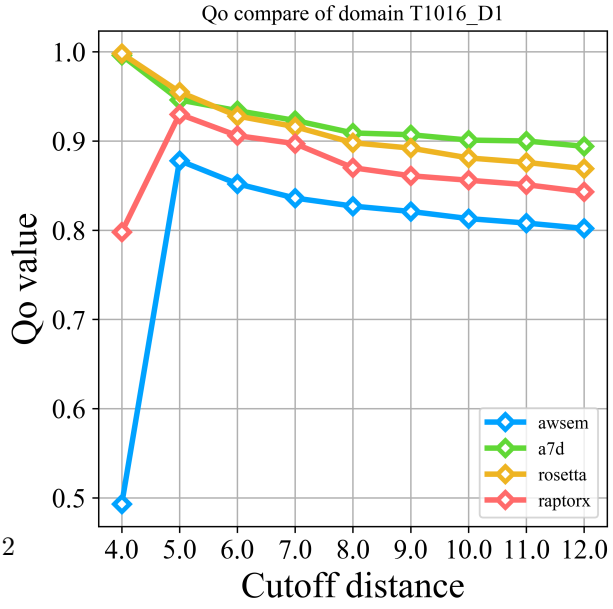

Qo compare of domain T1003_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

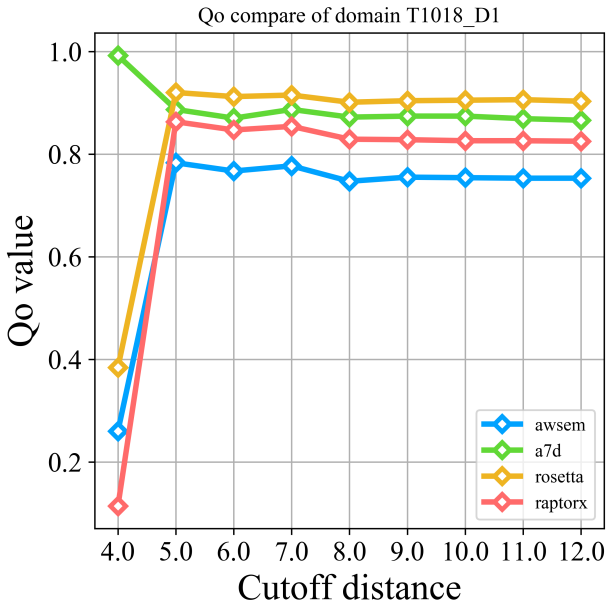

Qo compare of domain T1005_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

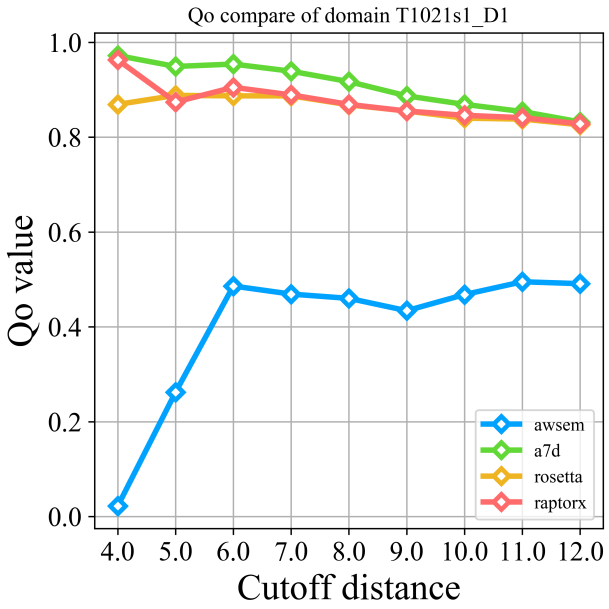

Qo compare of domain T1006_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

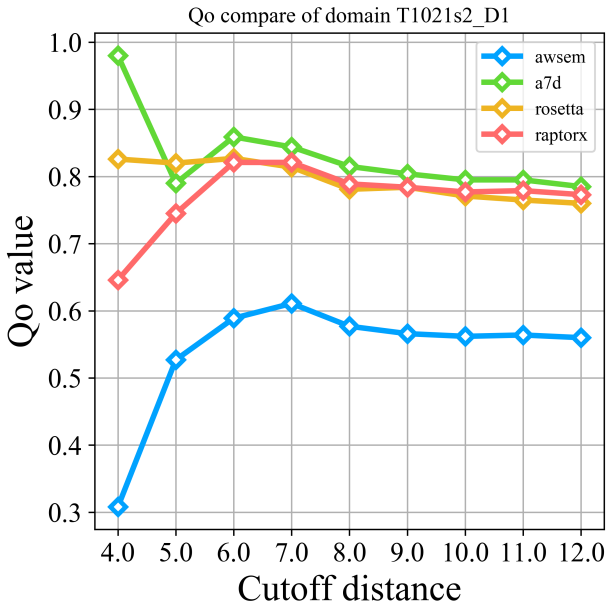

Qo compare of domain T1008_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

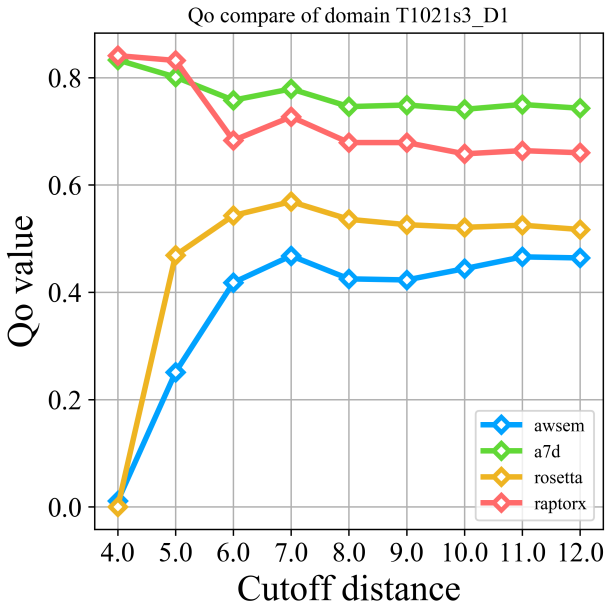

Qo compare of domain T1011_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

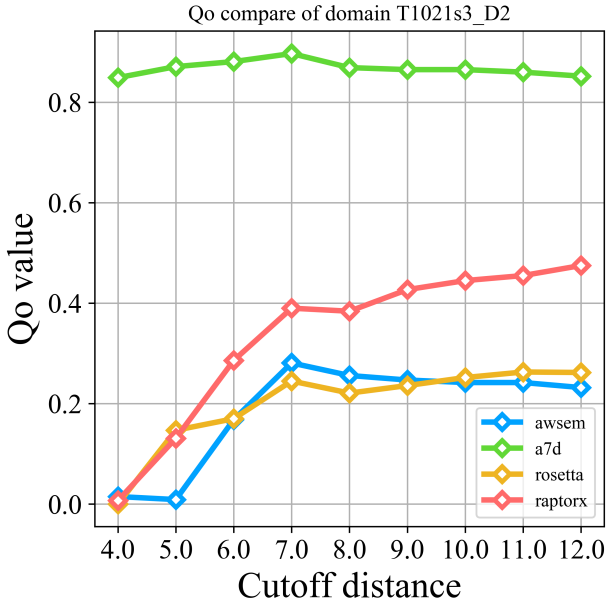

Qo compare of domain T1016_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

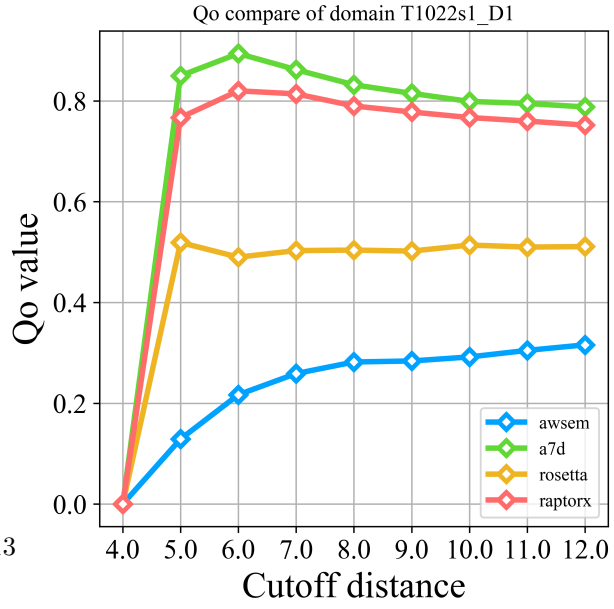

Qo compare of domain T1018_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

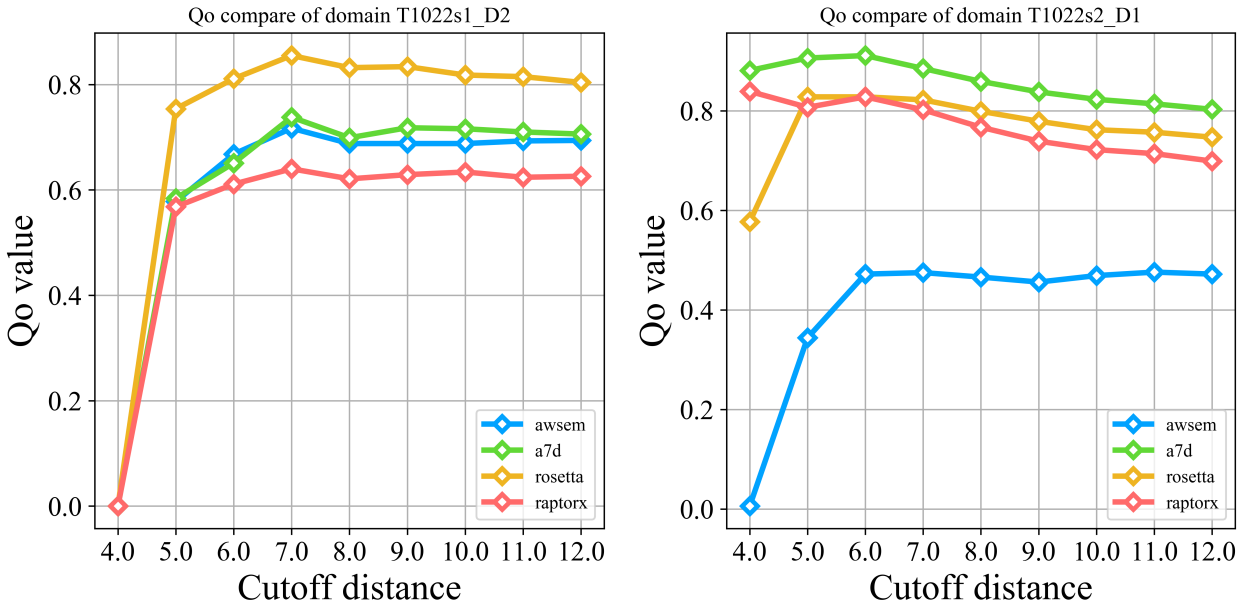

Qo compare of domain T1021s1_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx


Qo compare of domain T1021s2_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx


Qo compare of domain T1021s3_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx


Qo compare of domain T1021s3_D2
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx


Qo compare of domain T1022s1_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx


Qo compare of domain T1022s1_D2
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx
Qo compare of domain T1022s2_D1
Qo value
Cutoff distance
awsem
a7d
rosetta
raptorx

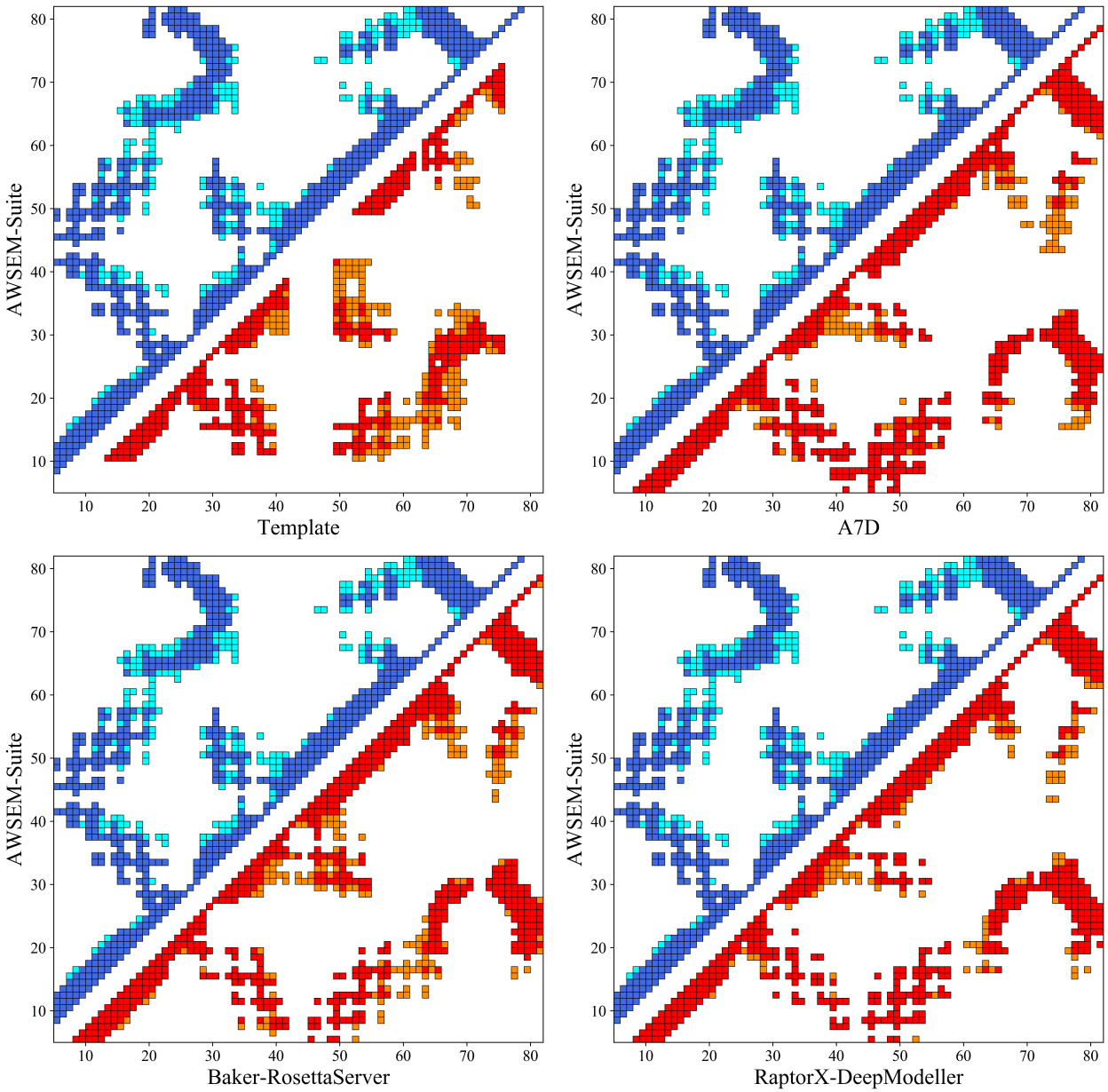


Figure S2: The contact map comparison of the AWSEM model with other predictor's structure and the template in T0958-D1. The cyan and orange squares represent the contact pairs that are not exist on crystal structure while blue and red squares represnt the correct contact pairs that occurs on crystal structure. The templates share an overall folding pattern with the crystal but lack of some important secondary structure information. AWSEM-Suite force field helps fold the correct tertiary structure in this case.

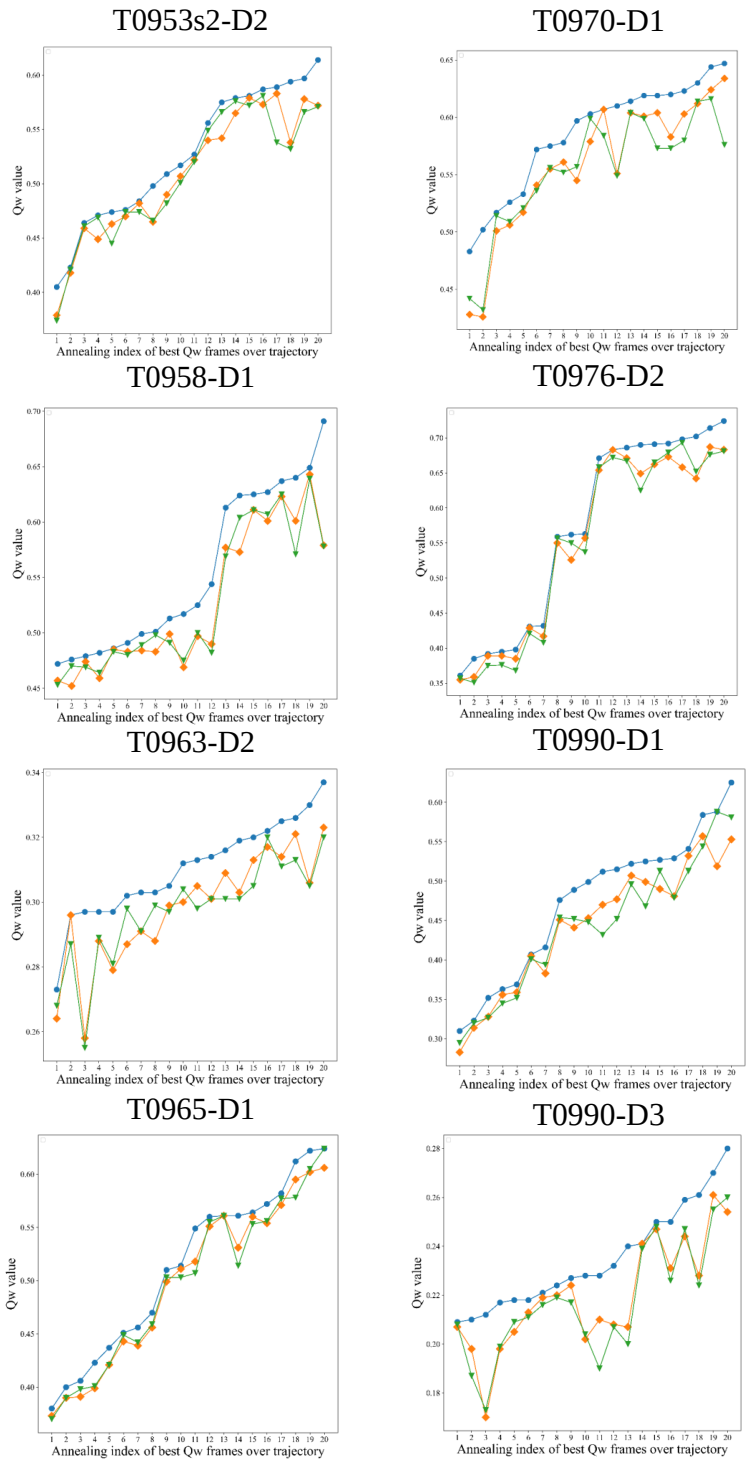


Figure S3: Summary of the quality of energy-based blind selection from prediction trajectory. In each case, the Q values of predicted structure from each of the 20 annealing simulations are plotted in ascending order based on the best Q frames over each trajectory. The Blue diamonds correspond to best Q frame of the trajectory. The orange diamonds correspond to best energy frame of the trajectory, which we picked for clustering and submitted in CASP13. The green diamonds correspond to the last frame of the trajectory.

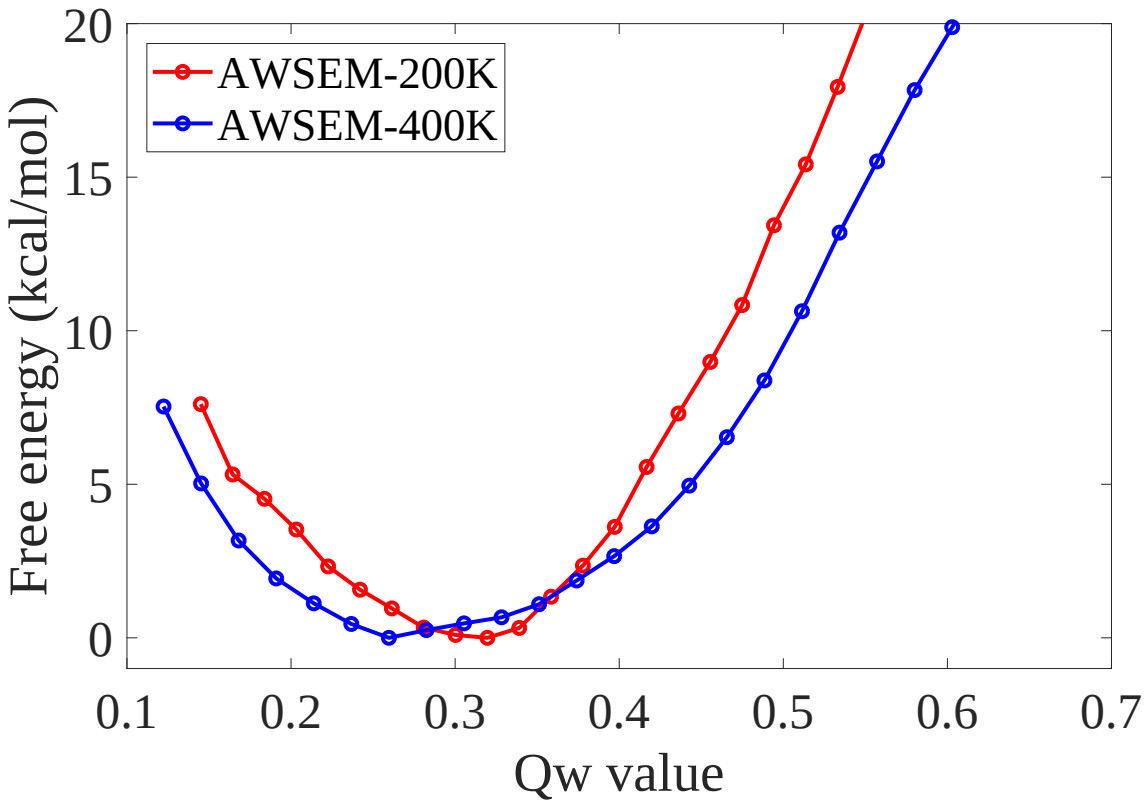


Figure S4: Free energy profiles of T0958 using AWSEM force field under 200 K and 400 K as a function of Q value. The performance of AWSEM force field have a slightly difference based on the natural property of sequence under different temperatures. Here the minimum free energy basin in both temperature have a lower Q value when compare with the simulation in 300 K.

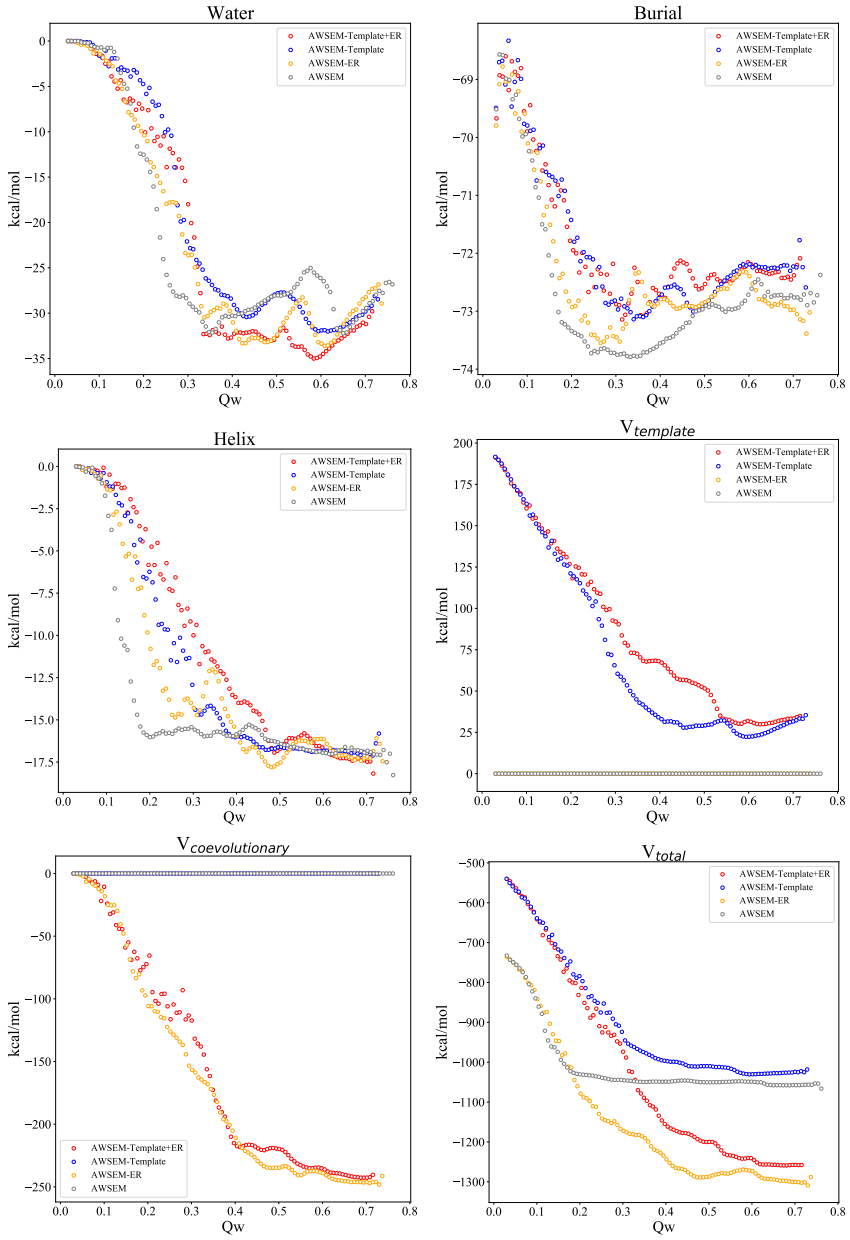


Figure S5: Values of six energy terms in different AWSEM models as a function of Q for the protein T0958. The energy terms in unit of kcal/mol are shown for several different models: AWSEM-Template+ER (red circles), AWSEM-Template (blue circles), AWSEM-ER (orange circles) and AWSEM (grey circles).

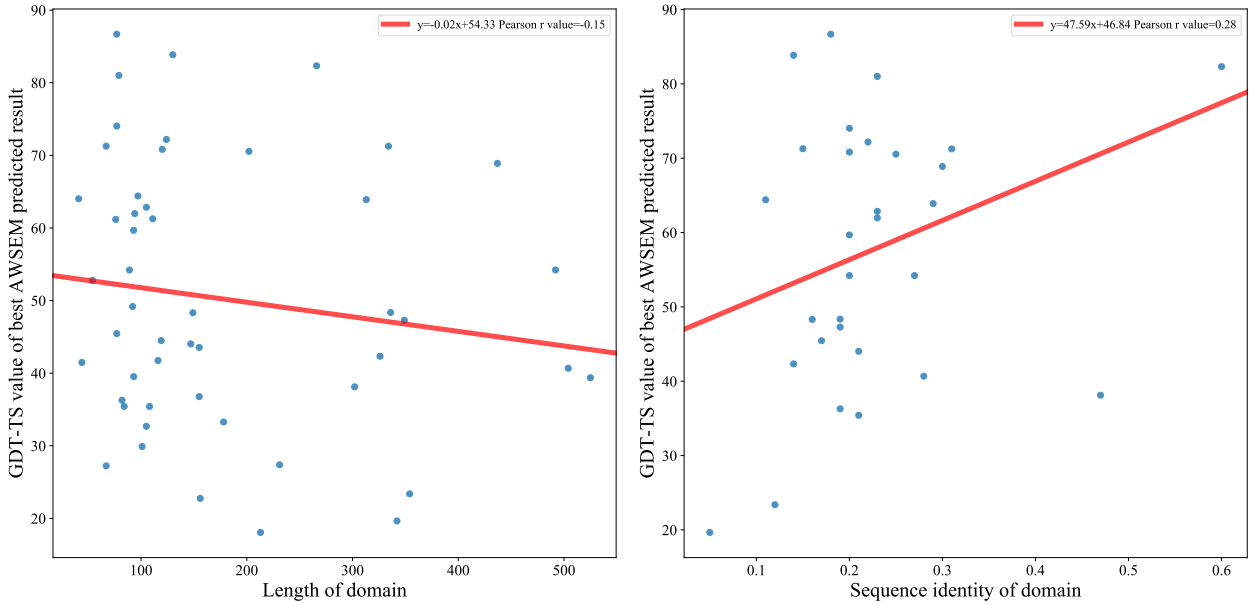


Figure S6: A) The correlation between the GDT-TS score and the length of each domain. B) The correlation between the GDT-TS score and sequence identity of template if available.

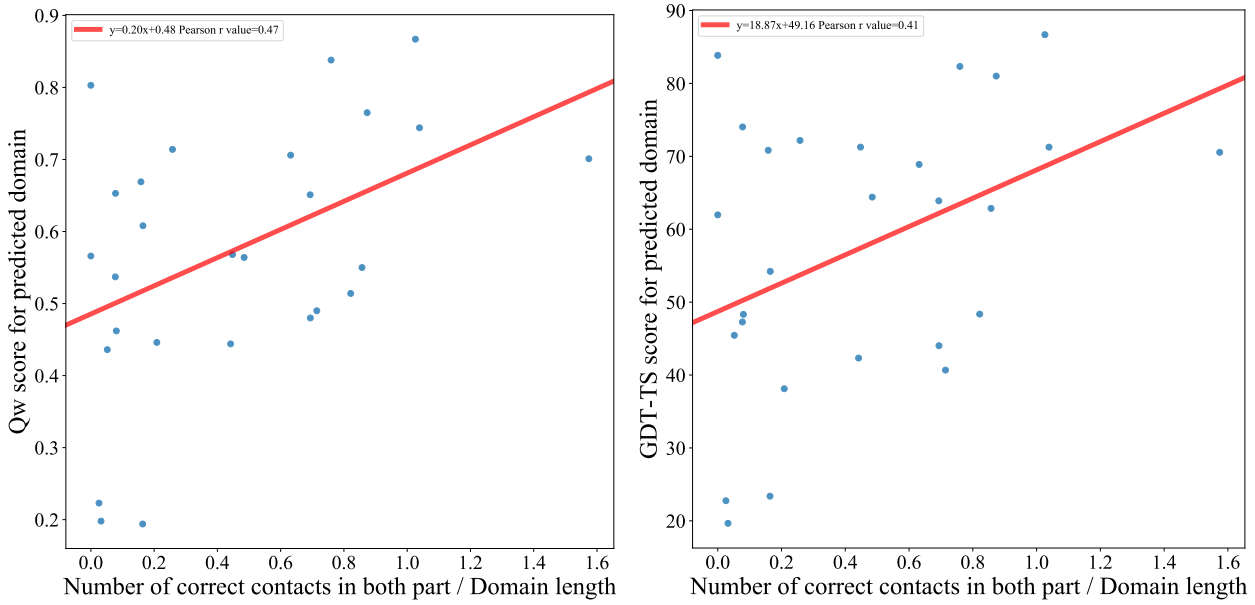


Figure S7: Ratio of correct contacts in both template and coevolutionary parts divided by domain length versus the Qw and GDT-TS value of each domain. The correct contacts in both template and coevolutionary parts and structure accuracy have a strong correlation(pearson = 0.47).

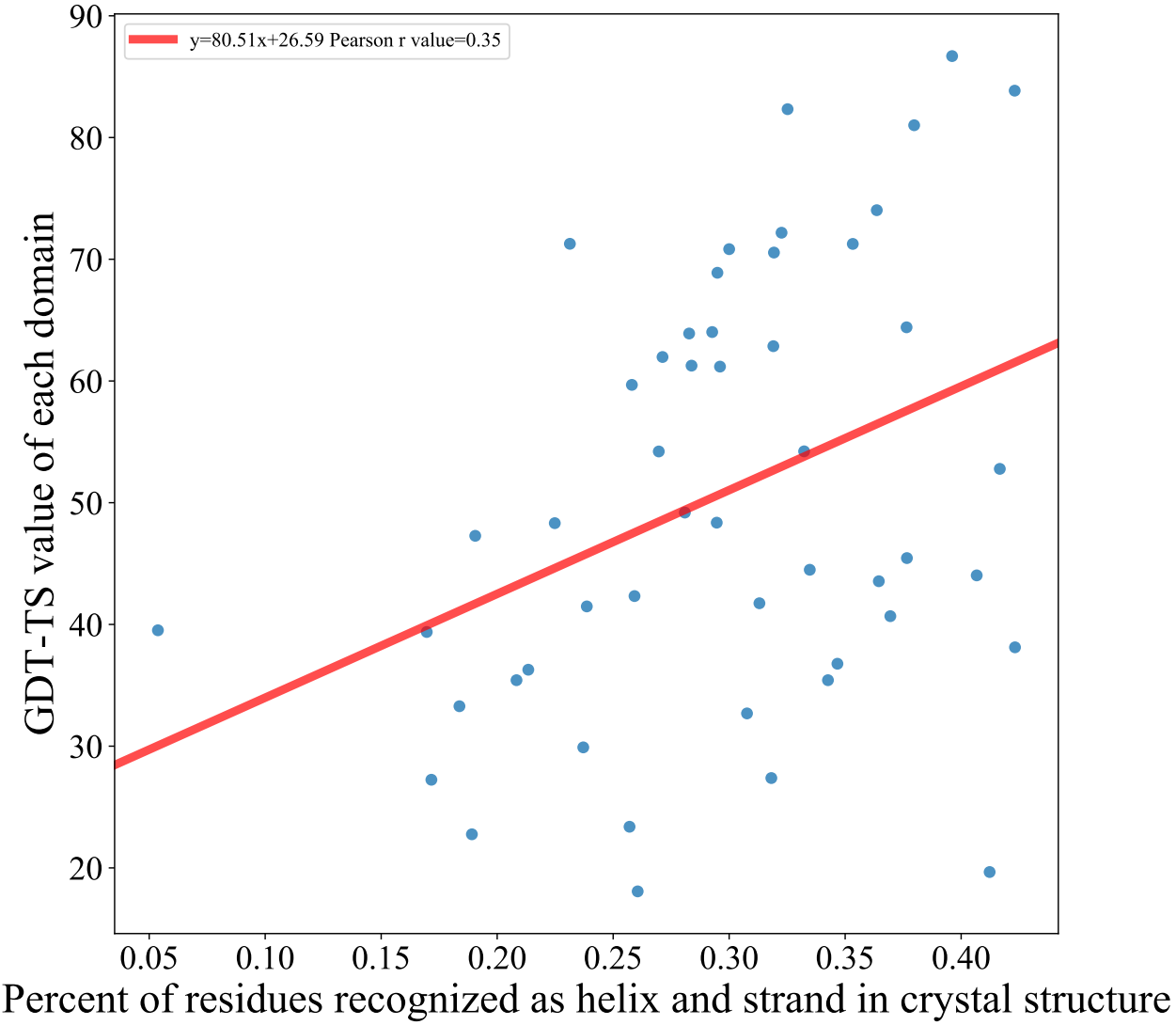


Figure S8: Qw value versus the helix plus strand ratio in the protein show the strong correlation between our predicted quality and the secondary structure.