

Statistical Analysis of EGFR Structures' Performance in Virtual Screening

Yan Li,^{,†,§} Xiang Li,^{‡,§} and Zigang Dong^{*,†}*

[†]The Hormel Institute, University of Minnesota, Austin Minnesota 55912, United States

[‡]Department of Physiology and Pathophysiology, School of Basic Medical Sciences, Zhengzhou University, 450001 Zhengzhou, Henan, China

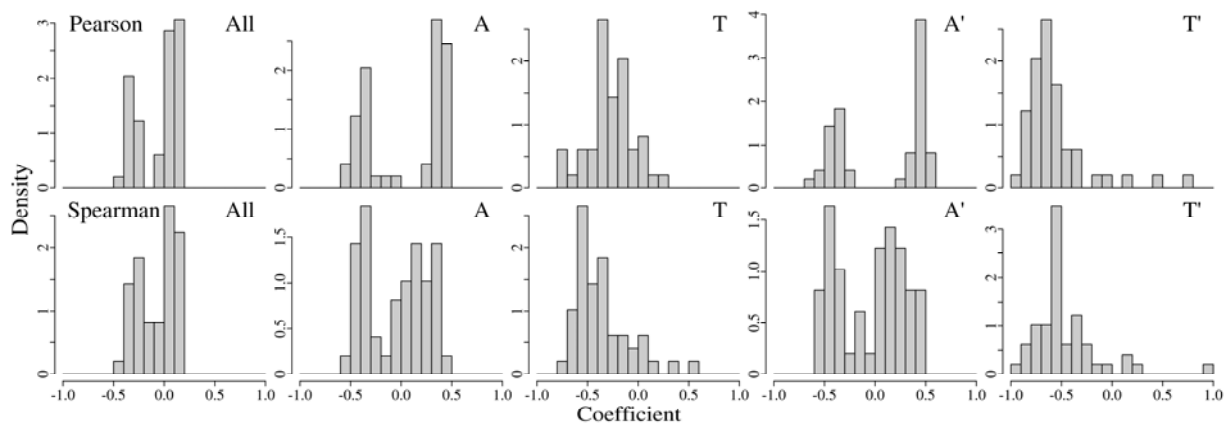


Figure S1. Distribution of coefficients between C α RMSD and EF with each EGFR structure as the template for five ensembles. The Pearson correlation coefficients are shown in the upper row. The Spearman correlation coefficients are displayed in the lower row.

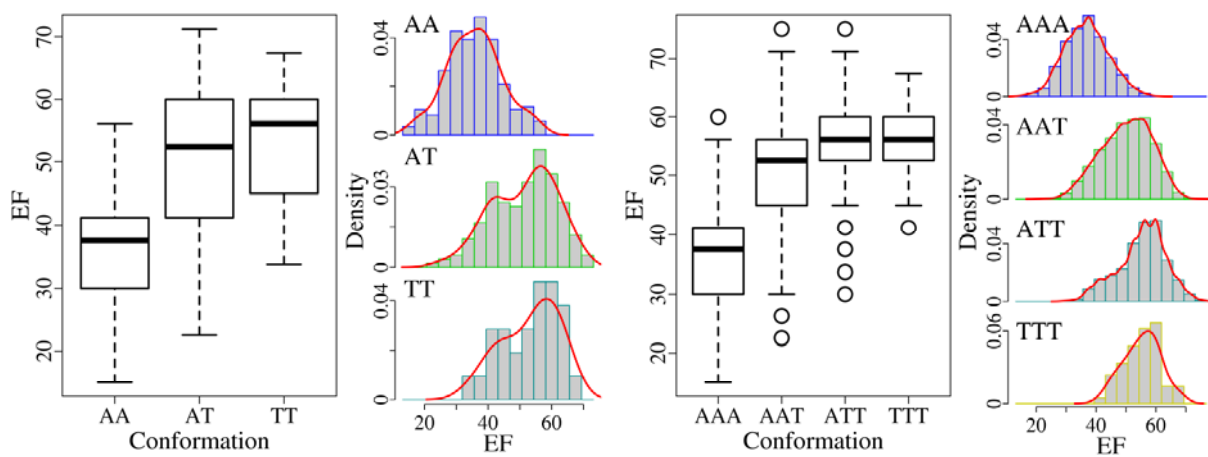
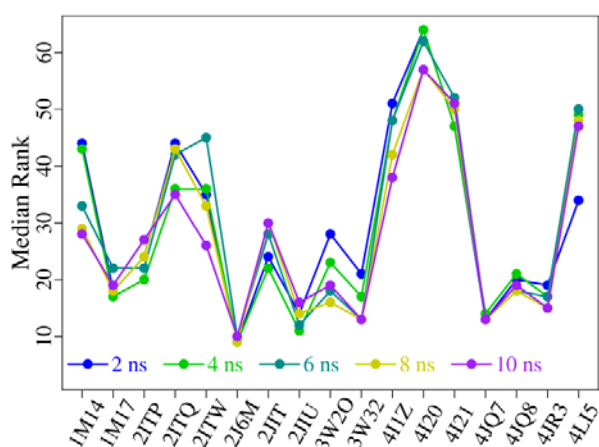
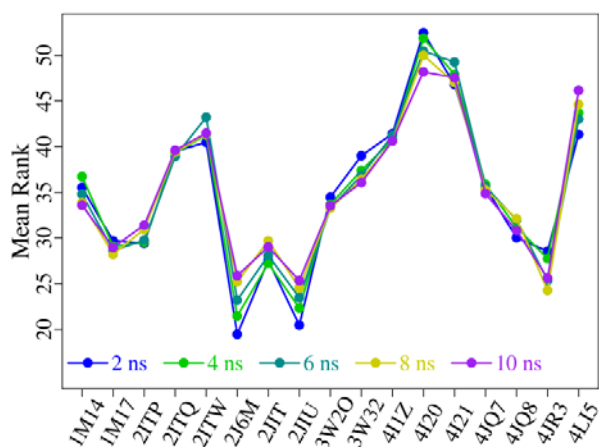
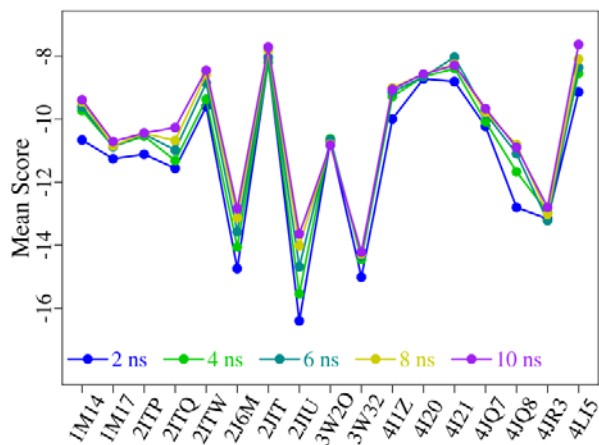


Figure S2. Discernment of two and three EGFR structures in virtual screening with the average score as the ligand score.

A



B

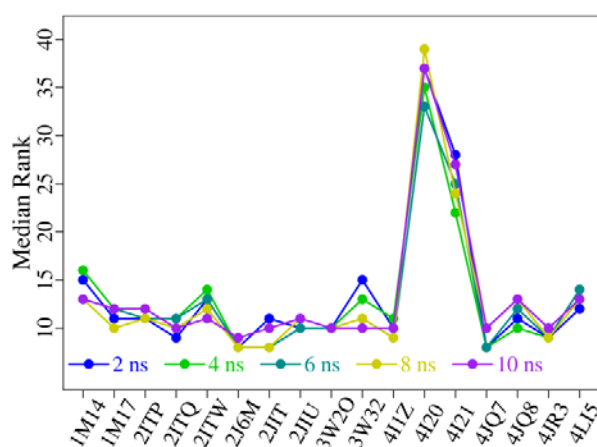
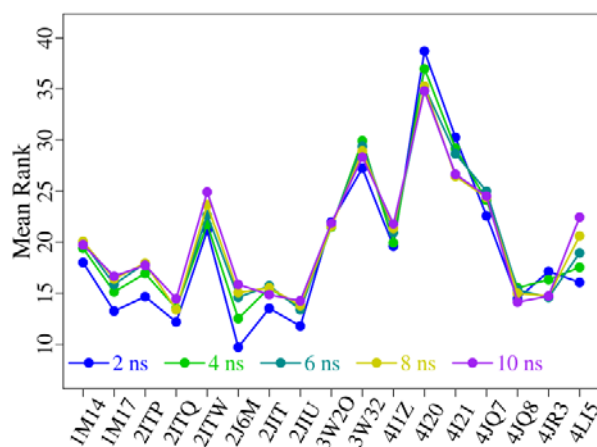
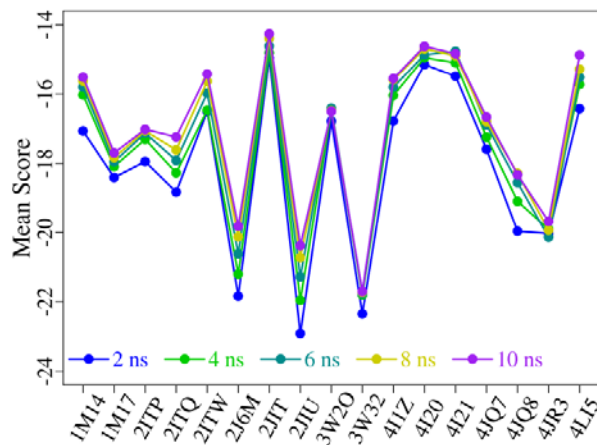


Figure S3. Effect of simulation length on the mean score, mean rank, and median rank of the 15 EGFR inhibitors in MM-PBSA. (A) The IDC is 2. (B) The IDC is 4.

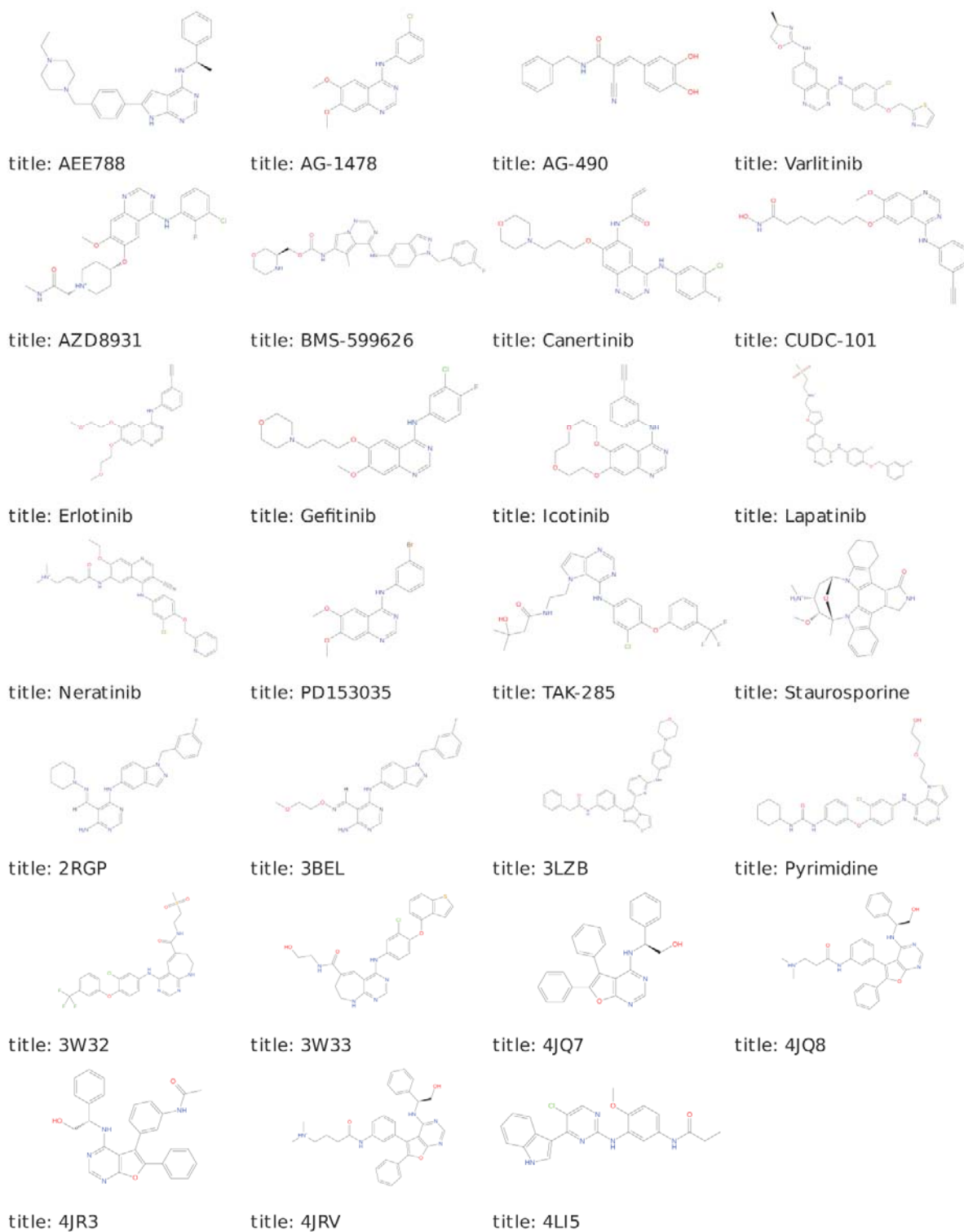


Figure S4. Chemical structures of 27 known EGFR inhibitors.

Table S1. The hypothesis testing results between two ensembles of EGFR structures calculated by T-test and Wilcoxon test as well as the mean EF with the standard deviation.

Ensemble 1			Ensemble 2			T-test			Wilcoxon test	
	size	EF		size	EF	p ¹	t	DF ²	p ¹	W
N	11	11.24±9.49	O	30	36.35±13.40	0	-6.671	25.314	0	20.5
N	11	11.24±9.49	P	8	21.07±8.72	0.033	-2.339	15.958	0.05	20
O	30	36.35±13.40	P	8	21.07±8.72	0.001	3.879	16.964	0.003	204
P	8	21.07±8.72	N'	7	14.99±9.44	0.22	1.291	12.385	0.265	38
A	33	28.05±11.75	T	12	36.53±20.55	0.198	-1.353	13.701	0.106	135
A	33	28.05±11.75	I	4	4.67±5.61	0	6.736	6.86	0.002	128
T	12	36.53±20.55	I	4	4.67±5.61	0	4.855	13.916	0.017	44
A'	22	32.19±11.49	T'	8	47.79±12.00	0.008	-3.185	12.005	0.007	31
W	22	32.87±14.86	M	27	24.42±16.05	0.062	1.91	46.187	0.102	378.5
W'	17	37.47±13.07	M'	13	34.88±14.22	0.613	0.513	24.753	0.949	108.5
WA ³	11	30.31±6.16	MA ⁴	11	34.06±15.22	0.461	-0.758	13.187	0.243	42.5
WT ⁵	6	50.60±12.28	MT ⁶	2	39.35±7.99	0.238	1.49	2.841	0.396	9

¹ p: p-value

² DF: degree of freedom

^{3,4} WA and MA: wild type (W) and mutated (M) EGFR structures adopting the active conformation (A) bound with an organic compound

^{5,6} WT and MT: wild type (W) and mutated (M) EGFR structures adopting the Src-like inactive conformation (T) bound with an organic compound

Table S2. The Pearson and Spearman correlation coefficients between the RMSD distance of EGFR structures and EF for five ensembles.

Ensemble	Size	Pearson				Spearman		
		r _p	p-value	t	DF	r _s	p-value	S
All	49	-0.045	0.757	-0.311	47	-0.077	0.6	21103.19
A	33	0.392	0.024	2.374	31	0.183	0.309	4890.827
T	12	-0.558	0.06	-2.125	10	-0.658	0.02	474.319
A'	22	0.437	0.042	2.17	20	0.055	0.809	1674.405
T'	8	-0.92	0.001	-5.763	6	-0.909	0.002	160.376

Table S3. The Pearson and Spearman correlation coefficients between the EF and ligand properties for three ensembles of EGFR structures.

Ensemble	Size	Property	Pearson				Spearman		
			r _p	p-value	t	DF	r _s	p-value	S
A'+T'	30	MW	0.119	0.53	0.636	28	0.063	0.743	4213.961
		AlogP	0.166	0.38	0.892	28	0.171	0.366	3725.198
		SA	0.174	0.356	0.938	28	0.134	0.481	3894.229
		HSA	0.389	0.033	2.236	28	0.276	0.14	3254.763
A'	22	MW	0.163	0.469	0.738	20	0.088	0.697	1615.411
		AlogP	0.329	0.135	1.557	20	0.225	0.313	1372.015
		SA	0.489	0.021	2.507	20	0.429	0.046	1010.516
		HAS	0.512	0.015	2.665	20	0.475	0.026	930.411
T'	8	MW	-0.184	0.664	-0.457	6	-0.279	0.504	107.422
		AlogP	-0.456	0.257	-1.254	6	-0.352	0.393	113.532
		SA	-0.253	0.545	-0.641	6	-0.376	0.359	115.569
		HSA	0.251	0.549	0.635	6	0.048	0.909	79.927

Table S4. The hypothesis testing results between ensembles of EGFR structures calculated from the lowest score and the average score as well as the mean EF with the standard deviation.

Ensemble	Size	Lowest	Average	T-test			Wilcoxon test	
		EF	EF	p	t	DF	p	W
AA	231	39.44±10.24	35.75±8.77	0	4.156	449.404	0	32296
AT	176	56.76±11.75	51.22±10.37	0	4.691	344.659	0	20034.5
TT	28	60.64±8.89	53.14±9.08	0.003	3.123	53.978	0.003	570
AAA	1540	43.96±9.85	36.78±7.33	0	22.939	2843.6	0	1680961.5
AAT	1848	60.33±11.02	50.15±8.45	0	31.533	3460.976	0	2644415.5
ATT	616	65.76±7.55	55.23±7.98	0	23.785	1226.305	0	319131.5
TTT	56	66.92±5.86	55.29±6.17	0	10.239	109.709	0	2854.5

Table S5. The hypothesis testing results in the two-structure and three-structure ensembles as well as the mean EF with the standard deviation.

Ensemble 1			Ensemble 2			T-test			Wilcoxon	
	size	EF		size	EF	p	t	DF	p	W
AA	231	39.44±10.24	AT	176	56.76±11.75	0	-15.564	347.535	0	5965
AA	231	39.44±10.24	TT	28	60.64±8.89	0	-11.709	36.257	0	437.5
AT	176	56.76±11.75	TT	28	60.64±8.89	0.047	-2.046	43.552	0.15	2048
AAA	1540	43.96±9.85	AAT	1848	60.33±11.02	0	-45.64	3369.265	0	420715
AAA	1540	43.96±9.85	ATT	616	65.76±7.55	0	-55.268	1465.352	0	47470
AAA	1540	43.96±9.85	TTT	56	66.92±5.86	0	-27.946	66.878	0	2501.5
AAT	1848	60.33±11.02	ATT	616	65.76±7.55	0	-13.648	1538.841	0	409755
AAT	1848	60.33±11.02	TTT	56	66.92±5.86	0	-8.005	67.411	0	33657
ATT	616	65.76±7.55	TTT	56	66.92±5.86	0.171	-1.383	72.752	0.4	16096.5

Table S6. The hypothesis testing results between the best one-structure, two-structure, and three-structure ensembles as well as the mean EF with the standard deviation.

Ensemble 1			Ensemble 2			T-test			Wilcoxon test	
	size	EF		size	EF	p	t	DF	p	W
T'	8	47.79±12.00	AT	176	56.76±11.75	0.074	-2.07	7.623	0.039	401
T'	8	47.79±12.00	TT	28	60.64±8.89	0.019	-2.818	9.312	0.01	44
T'	8	47.79±12.00	ATT	616	65.76±7.55	0.004	-4.228	7.072	0	443
T'	8	47.79±12.00	TTT	56	66.92±5.86	0.003	-4.437	7.484	0	31
AT	176	56.76±11.75	ATT	616	65.76±7.55	0	-9.62	217.92	0	29328
AT	176	56.76±11.75	TTT	56	66.92±5.86	0	-8.605	188.781	0	2333
TT	28	60.64±8.89	ATT	616	65.76±7.55	0.006	-2.998	28.798	0.002	5634.5
TT	28	60.64±8.89	TTT	56	66.92±5.86	0.002	-3.388	39.07	0.001	447

Table S7. The hypothesis testing results between ensembles of the active structures as well as the mean EF with the standard deviation.

Ensemble 1			Ensemble 2			T-test			Wilcoxon test	
	size	EF		size	EF	p	t	DF	p	W
A	22	32.19±11.49	AA	231	39.44±10.24	0.009	-2.855	24.282	0.003	1589.5
A	22	32.19±11.49	AAA	1540	43.96±9.85	0	-4.781	21.443	0	7034.5
AA	231	39.44±10.24	AAA	1540	43.96±9.85	0	-6.287	297.481	0	134129

Table S8. The mean Glide score of the top 1%, 2%, 10%, and all compounds with the standard deviation.

Structure	1%	2%	10%	All
1M14	-7.699±0.302	-7.470±0.315	-6.884±0.365	-5.396±0.888
1M17	-8.255±0.536	-7.979±0.469	-7.361±0.407	-5.809±0.937
1XKK	-8.923±1.107	-8.444±0.915	-7.671±0.586	-6.051±0.982
2EB2	-8.022±0.265	-7.810±0.285	-7.244±0.351	-5.629±0.974
2EB3	-8.029±0.278	-7.816±0.293	-7.249±0.352	-5.643±0.970
2GS2	-7.576±0.277	-7.368±0.288	-6.805±0.349	-5.315±0.890
2GS7	-8.210±0.338	-7.976±0.337	-7.387±0.368	-5.775±0.970
2ITN	-8.141±0.253	-7.936±0.275	-7.365±0.349	-5.784±0.968
2ITO	-8.119±0.570	-7.842±0.489	-7.219±0.411	-5.718±0.907
2ITP	-8.106±0.460	-7.856±0.410	-7.258±0.389	-5.730±0.915
2ITQ	-8.348±0.295	-8.126±0.308	-7.564±0.354	-5.927±1.003
2ITT	-8.207±0.634	-7.910±0.537	-7.229±0.447	-5.670±0.925
2ITU	-8.061±0.219	-7.867±0.252	-7.317±0.339	-5.644±1.026
2ITV	-7.902±0.223	-7.716±0.246	-7.198±0.327	-5.620±0.955
2ITW	-7.902±0.207	-7.726±0.233	-7.202±0.321	-5.642±0.947
2ITX	-8.148±0.262	-7.945±0.277	-7.390±0.347	-5.751±0.985
2ITY	-8.000±0.567	-7.711±0.494	-7.090±0.412	-5.598±0.894
2ITZ	-8.257±0.554	-7.966±0.487	-7.325±0.426	-5.751±0.957
2J6M	-8.074±0.497	-7.852±0.415	-7.277±0.379	-5.752±0.913
2JIT	-6.653±0.191	-6.492±0.213	-6.064±0.264	-4.926±0.759
2JIU	-8.488±0.694	-8.169±0.584	-7.486±0.460	-5.881±0.956
2RF9	-6.576±0.196	-6.410±0.219	-5.961±0.274	-4.827±0.722
2RFD	-8.048±0.248	-7.841±0.273	-7.296±0.339	-5.689±0.977
2RFE	-7.968±0.266	-7.745±0.296	-7.148±0.362	-5.632±0.909
2RGP	-8.861±0.752	-8.486±0.650	-7.772±0.488	-6.155±0.991
3BEL	-8.948±1.125	-8.500±0.909	-7.729±0.582	-6.127±0.982
3GT8	-8.584±0.393	-8.309±0.392	-7.654±0.411	-5.999±1.002
3LZB	-8.446±0.515	-8.135±0.479	-7.488±0.414	-6.065±0.894
3POZ	-8.918±1.064	-8.547±0.836	-7.848±0.534	-6.221±1.019
3UG1	-7.992±0.233	-7.793±0.263	-7.144±0.397	-5.509±0.969
3UG2	-7.987±0.255	-7.790±0.268	-7.262±0.330	-5.656±0.980
3VJN	-7.789±0.273	-7.577±0.290	-7.034±0.334	-5.505±0.941
3VJO	-7.747±0.264	-7.518±0.299	-6.924±0.367	-5.326±0.945
3W2O	-8.150±0.246	-7.953±0.266	-7.387±0.352	-5.724±1.008
3W2R	-8.655±0.337	-8.398±0.352	-7.773±0.388	-6.039±1.050
3W2S	-8.718±0.356	-8.455±0.369	-7.774±0.418	-6.070±1.026
3W32	-9.184±1.106	-8.812±0.862	-8.096±0.553	-6.460±1.027
3W33	-8.891±1.190	-8.504±0.922	-7.794±0.567	-6.176±1.000
4HJO	-8.741±0.393	-8.457±0.399	-7.800±0.414	-6.045±1.063
4IIZ	-7.889±0.288	-7.669±0.303	-7.070±0.370	-5.547±0.911
4I20	-7.166±0.348	-6.925±0.348	-6.320±0.375	-5.064±0.769

4I21	-8.332±0.305	-8.048±0.361	-7.304±0.456	-5.545±1.000
4I22	-8.947±0.530	-8.659±0.473	-7.997±0.428	-6.249±1.065
4I23	-8.349±0.340	-8.115±0.338	-7.489±0.388	-5.854±0.978
4JQ7	-8.157±0.588	-7.877±0.502	-7.237±0.427	-5.732±0.900
4JQ8	-8.352±0.643	-8.050±0.545	-7.397±0.438	-5.766±0.968
4JR3	-8.112±0.545	-7.817±0.485	-7.145±0.432	-5.610±0.905
4JRV	-8.300±0.636	-8.040±0.518	-7.419±0.424	-5.758±0.998
4LI5	-8.336±0.292	-8.117±0.304	-7.524±0.370	-5.761±1.059