

# Supporting Information: MELD x MD Folds Nonthreadables, Giving Native Structures and Populations

James C. Robertson,<sup>†</sup> Alberto Perez,<sup>†</sup> and Ken A. Dill<sup>\*,†,‡,¶</sup>

<sup>†</sup>*Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony  
Brook, New York 11794*

<sup>‡</sup>*Department of Chemistry, Stony Brook University, Stony Brook, New York 11794*

<sup>¶</sup>*Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York  
11794*

E-mail: [dill@laufercenter.org](mailto:dill@laufercenter.org)

## Supporting Information

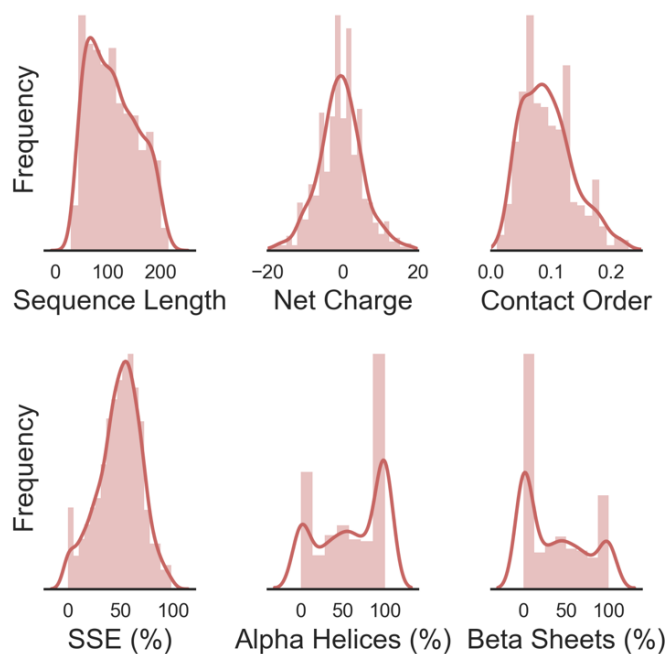


Figure 1: **Characteristics of nonthreadable proteins.** Histograms of nonthreadable properties are shown with a (solid line) smoothed estimate of the distribution from kernel density estimation. The sequence length, net charge, relative contact order, and secondary structure elements (SSE) were calculated for 893 nonthreadable proteins (five proteins over 215 amino acids were excluded).

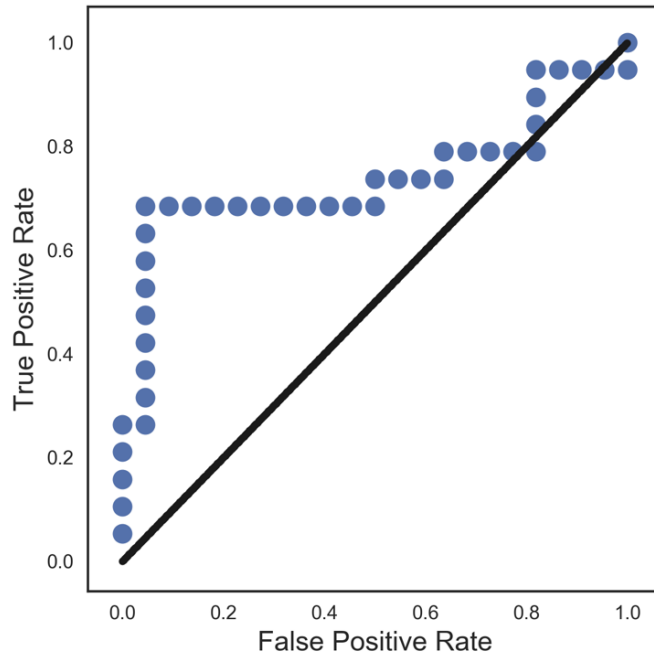


Figure 2: **MELD x MD populations provide measure of confidence for prediction.** High MELD x MD populations were observed in simulations that folded to native, while low populations often folded to non-native. The true positive versus false positive rates for 41 nonthreadables are shown in the receiver operator characteristic (ROC) curve. The area under the ROC was 0.74.

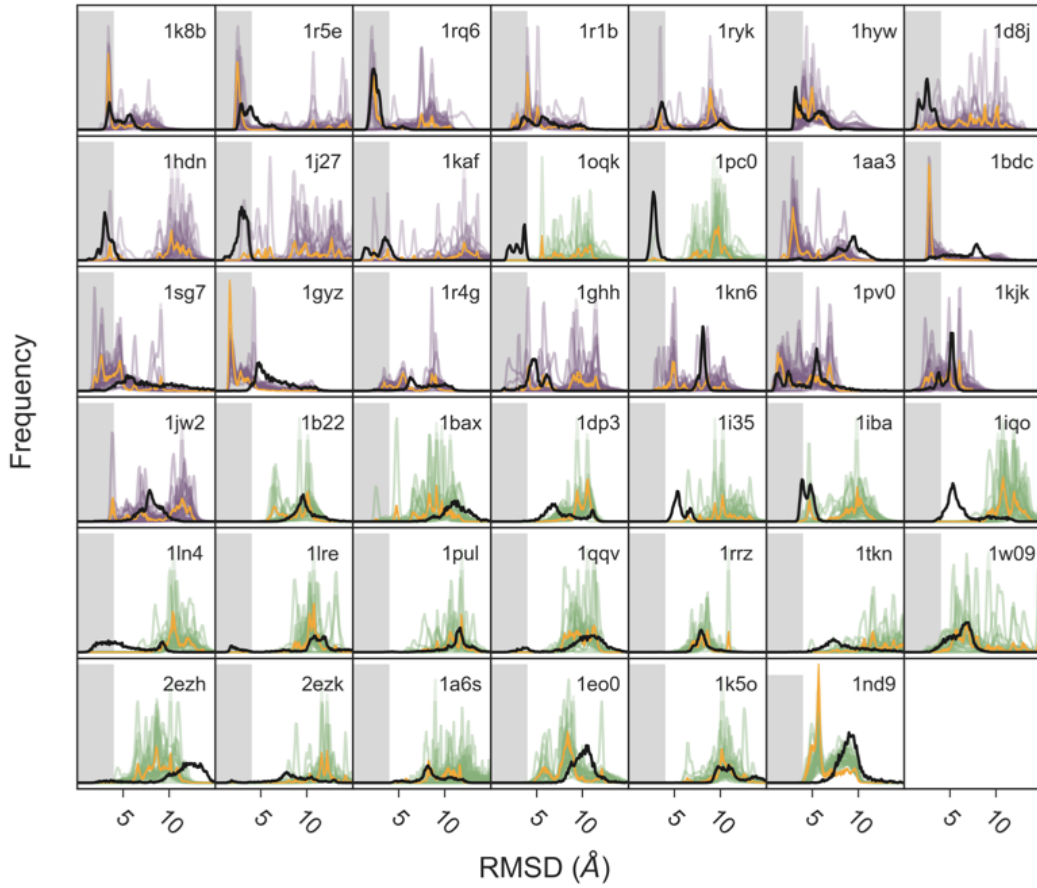


Figure 3: **MELD x MD and MD stability tests do not always agree on native and non-native.** There are three sets of RMSD-to-native histograms for 41 nonthreadables: (black) MD stability tests, (orange) the lowest five temperature replicas from MELD x MD, all thirty replicas followed through the replica exchange ladder of MELD x MD (purple) folders and (green) nonfolders. The shaded region in each is from 0 to 4 Å. Note that MD stability tests were started from native conformation, MELD x MD runs were started from extended.

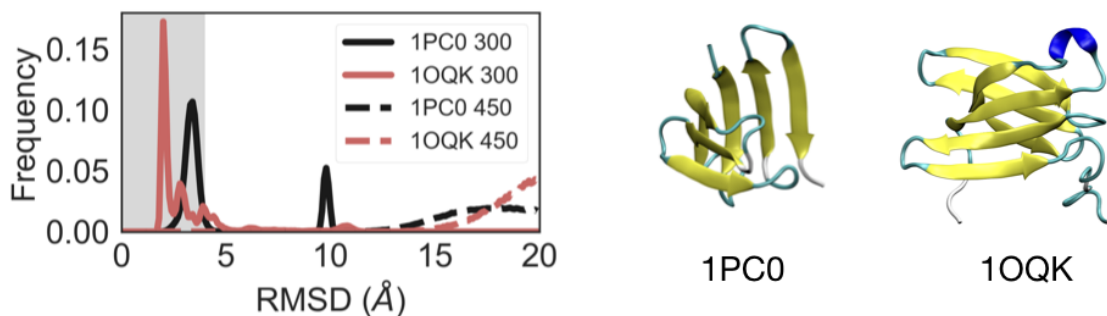
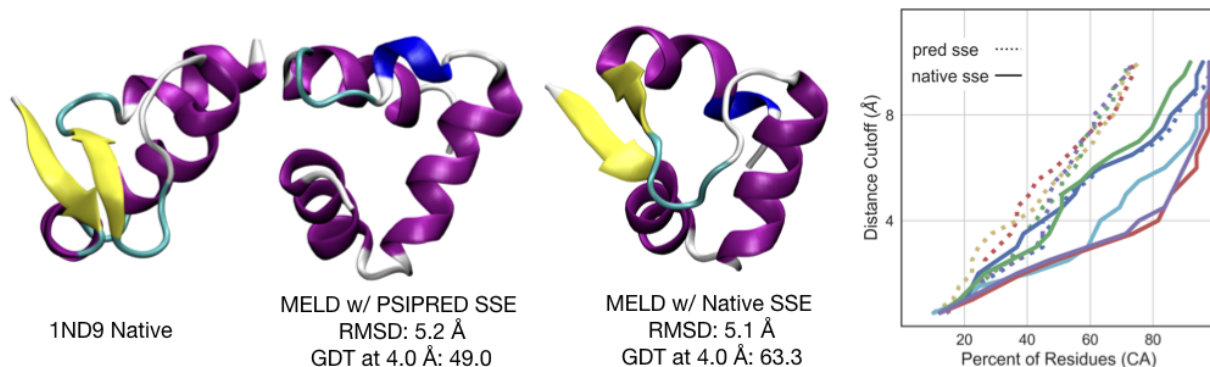


Figure 4: **MELD x MD does not fold two force-field-stable all beta nonthreadables.** MELD x MD did not fold these proteins to native. The MD stability tests indicated they were stable in the force field but we wondered if they were trapped in the native state at low temperature. Hence, we ran 1PC0 and 1OQK in T-REMD stability tests. The RMSD-to-native plot on the left shows both have low RMSD for the 300 K replica, which indicates that the two proteins are stable in the force field and solvent model. The highest temperature replica (450 K) RMSDs are also shown to demonstrate that the protein does unfold at high temperatures. The structure of the two proteins are shown to the right with (yellow) beta sheets, (cyan) turns, (white) loops, and (blue) 3<sub>10</sub> helices.



TDVTIKTLAAERQTSVERLVQQFADAGIRKSADDSVSAQEKQTLIDHLN	Sequence
...HHHHHHHHHH...HHHHHHHHHHHH.....HHHHHHHHHHHH.	PSIPRED
..HHHHHHHHHHHH...HHHHHHHHHHHH.....HHHHHHHHHHHH.	MELD w/ PSIPRED
.EE...HHHHHH...HHHHHHHHHHHH.....EE.....HHHHHH.	Native
.EEEHHHHHHHH...HHHHHHHHHHHH.....EEEHHHHHH.HHHH.	MELD w/ Native

Figure 5: **Inaccurate information and force field deficiencies sometimes limit MELD x MD.** 1ND9 structures: the PDB reference, MELD x MD simulated with PSIPRED predicted sse, and MELD x MD simulated with native sse (colored yellow for beta sheets, purple for alpha helices, cyan for turns, white for coils, and blue for 3<sub>10</sub> helices). The plot shows GDT scores for the top 5 MELD x MD structures simulated with PSIPRED secondary structure (dashed lines) compared to native secondary structure (solid lines). The GDT colors from highest to lowest population are teal, blue, green, red, purple. SSEs are aligned to the sequence in the following order: PSIPRED predicted sse, MELD x MD given PSIPRED, native, and MELD x MD given native sse. This shows that bad sse information limits MELD x MD, bet even when corrected the force field can discourage MELD x MD from folding to native.

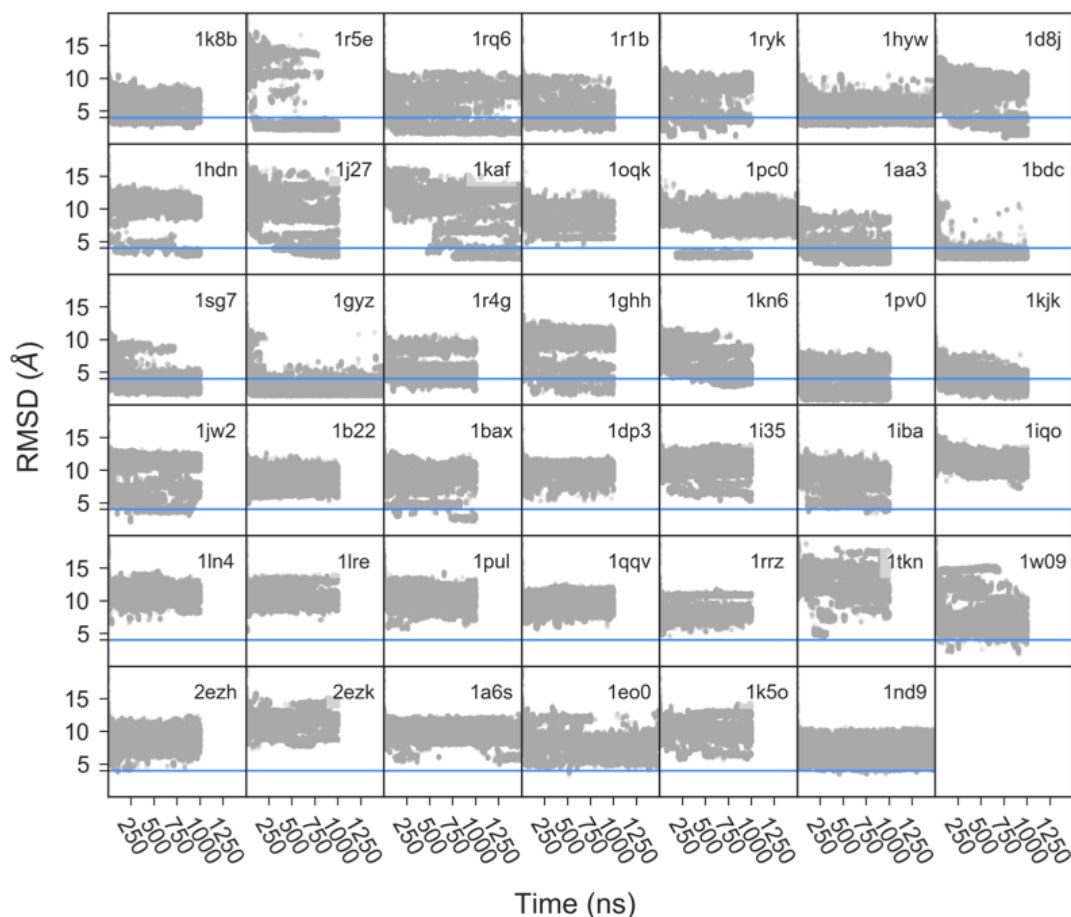


Figure 6: **MELD x MD folds many nonthreadables to native with less than 1  $\mu$ s sampling.** The RMSD vs. time series are shown for 41 nonthreadables. Many that fold to native do so within 250 ns, while others like 1KAF require more than 1  $\mu$ s. 1PC0 appears to fold to native, but with sampling extended to 1.5  $\mu$ s, the native structure is no longer populated. Others, like 1BAX, may fold with more sampling. RMSD was calculated for  $C\alpha$  and  $C\beta$  atoms of residues predicted to have secondary structure elements. The RMSD data are for the five lowest temperature replicas, the same replicas that were clustered to determine folded or nonfolded. The blue line in each plot is at 4.0 Å, the cutoff for folded to native.

Table 1: MELD x MD Candidates with sequence length, RMSD to PDB reference and cluster populations for the top three clusters (c0-c2). MELD x MD folders (top half of table) have PDB ID, RMSD, and cluster populations in bold. The amount of native secondary structure is also shown.

PDB ID	Seq. Length	RMSD (Å)			Population			Sec. Struct. (%)		
		c0	c1	c2	c0	c1	c2	Total	Alpha	Beta
<b>1aa3</b>	63	<b>2.8</b>	2.9	5.6	<b>0.62</b>	0.11	0.08	0.43	0.78	0.22
<b>1bdc</b>	60	<b>2.8</b>	3.9	4.2	<b>0.96</b>	0.04	0.00	0.55	1.00	0.00
<b>1d8j</b>	81	7.3	<b>2.3</b>	8.8	0.16	<b>0.15</b>	0.13	0.54	0.84	0.16
<b>1ghh</b>	81	<b>4.0</b>	11.4	5.9	<b>0.13</b>	0.12	0.12	0.68	0.65	0.35
<b>1hdn</b>	85	10.4	10.0	<b>3.5</b>	0.13	0.09	<b>0.09</b>	0.67	0.56	0.44
<b>1j27</b>	98	4.5	<b>3.2</b>	5.9	0.14	<b>0.10</b>	0.09	0.76	0.47	0.53
<b>1jw2</b>	72	<b>4.0</b>	10.8	11.4	<b>0.20</b>	0.08	0.08	0.67	1.00	0.00
<b>1k8b</b>	52	<b>3.4</b>	5.4	5.8	<b>0.73</b>	0.03	0.02	0.75	0.54	0.46
<b>1kjk</b>	49	<b>2.8</b>	6.1	5.5	<b>0.56</b>	0.08	0.06	0.57	0.61	0.39
<b>1kn6</b>	73	<b>3.4</b>	8.6	4.7	<b>0.15</b>	0.15	0.14	0.42	0.55	0.45
<b>1pv0</b>	46	<b>1.1</b>	6.8	5.1	<b>0.51</b>	0.21	0.12	0.59	1.00	0.00
<b>1r1b</b>	56	<b>4.0</b>	7.4	3.7	<b>0.64</b>	0.05	0.05	0.61	1.00	0.00
<b>1r4g</b>	53	5.2	<b>3.3</b>	8.5	0.42	<b>0.17</b>	0.12	0.68	1.00	0.00
<b>1r5e</b>	105	<b>2.4</b>	10.8	14.0	<b>0.85</b>	0.03	0.03	0.65	1.00	0.00
<b>1ryk</b>	69	8.9	<b>3.5</b>	8.2	0.41	<b>0.12</b>	0.07	0.65	1.00	0.00
<b>1sg7</b>	75	<b>3.5</b>	4.7	9.3	<b>0.78</b>	0.06	0.05	0.60	0.87	0.13
<b>1gyz</b>	60	<b>1.6</b>	4.2	3.7	<b>0.85</b>	0.03	0.02	0.63	1.00	0.00
<b>1hyw</b>	58	<b>3.1</b>	5.7	5.0	<b>0.40</b>	0.09	0.09	0.62	0.78	0.22
<b>1kaf</b>	108	13.6	6.6	<b>2.5</b>	0.08	0.08	<b>0.07</b>	0.68	0.51	0.49
<b>1rq6</b>	62	<b>2.1</b>	8.4	4.5	<b>0.65</b>	0.09	0.06	0.63	1.00	0.00
1b22	70	6.7	10.1	9.3	0.17	0.10	0.10	0.47	0.94	0.06
1bax	94	9.1	8.3	9.6	0.16	0.13	0.10	0.61	1.00	0.00
1dp3	55	9.4	10.6	9.4	0.19	0.19	0.12	0.55	1.00	0.00
1i35	95	10.3	10.2	9.3	0.15	0.10	0.10	0.52	0.49	0.51
1iba	78	4.6	10.2	9.9	0.11	0.10	0.09	0.53	0.68	0.32
1iqo	88	10.8	9.2	10.5	0.12	0.11	0.07	0.41	0.61	0.39
1ln4	98	10.5	12.4	10.0	0.15	0.11	0.10	0.60	0.68	0.32
1lre	81	10.9	10.3	10.7	0.17	0.16	0.12	0.68	1.00	0.00
1oqk	78	5.6	10.8	9.6	0.14	0.11	0.08	0.44	0.00	1.00
1pul	103	11.8	10.8	7.7	0.17	0.08	0.07	0.67	1.00	0.00
1qqv	67	11.3	8.0	10.5	0.12	0.11	0.10	0.42	0.93	0.07
1rrz	66	8.0	10.9	7.1	0.28	0.15	0.08	0.70	1.00	0.00
1tkn	110	11.7	11.4	10.9	0.15	0.11	0.07	0.91	1.00	0.00
1w09	92	6.6	5.4	7.1	0.18	0.14	0.11	0.78	1.00	0.00
2ezh	65	8.8	8.7	10.2	0.10	0.09	0.08	0.69	1.00	0.00
2ezk	93	12.2	11.8	11.7	0.17	0.13	0.09	0.69	1.00	0.00
1pc0	61	9.5	9.2	10.9	0.16	0.07	0.06	0.52	0.00	1.00
1a6s	87	8.3	11.7	8.7	0.15	0.14	0.09	0.53	1.00	0.00
1eo0	77	5.3	8.1	8.5	0.19	0.15	0.06	0.69	1.00	0.00
1k5o	86	10.2	12.6	9.8	0.16	0.09	0.09	0.49	1.00	0.00
1nd9	49	5.2	9.4	8.5	0.66	0.10	0.02	0.55	0.85	0.15



Table 2: MELD x MD candidates that never sampled native were re-run with MELD x MD starting from the native conformation. These seeded MELD x MD simulations were run to determine if sampling was limiting MELD x MD from populating the native state. Only one of 15 starting from native had a low free energy structure close to native, 1LN4, suggesting that lack of sampling was not responsible for non-native MELD x MD structures. The RMSD to PDB reference and cluster populations for the top five clusters (c0-c4) are shown, PDB ID, RMSD, and cluster populations in bold for 1LN4.

PDB ID	Seq. Length	RMSD (Å)					Population				
		c0	c1	c2	c3	c4	c0	c1	c2	c3	c4
1b22	70	11.0	9.5	10.3	10.4	10.1	0.15	0.11	0.08	0.08	0.07
1dp3	55	9.6	10.3	10.4	10.9	10.6	0.11	0.09	0.09	0.06	0.05
1i35	95	8.4	12.4	8.9	10.1	11.8	0.12	0.10	0.08	0.07	0.05
1iqo	88	13.1	13.2	9.5	10.1	10.6	0.15	0.10	0.09	0.08	0.06
<b>1ln4</b>	98	<b>1.2</b>	11.0	10.4	11.4	8.7	<b>0.21</b>	0.19	0.19	0.11	0.10
1lre	81	8.9	10.3	11.4	9.3	9.7	0.34	0.19	0.11	0.05	0.04
1oqk	78	11.0	10.0	9.9	9.6	10.6	0.10	0.08	0.08	0.07	0.07
1pul	103	11.6	10.8	11.1	11.8	8.3	0.14	0.13	0.13	0.09	0.06
1qqv	67	11.4	9.0	10.6	9.1	10.6	0.08	0.07	0.06	0.06	0.05
1rrz	66	8.4	7.2	7.2	8.2	7.2	0.26	0.12	0.09	0.07	0.05
1tkn	110	11.0	9.7	9.9	7.7	8.6	0.14	0.12	0.09	0.07	0.07
2ezh	65	8.6	6.3	7.3	11.4	8.7	0.33	0.19	0.05	0.04	0.04
2ezk	93	4.6	14.7	12.2	10.1	8.9	0.37	0.15	0.11	0.08	0.03
1a6s	87	10.8	11.0	10.2	11.2	7.8	0.13	0.13	0.11	0.08	0.04
1k5o	86	9.7	8.7	10.1	10.1	10.3	0.11	0.10	0.09	0.07	0.06

Table 3: The relative contact order for 41 nonthreadables was calculated from the PDB reference structure. MELD x MD folded high and low contact order proteins. MELD x MD folded protein (MELD x MD result 0), MELD x MD did not fold protein (MELD x MD result 1).

PDB ID	Seq. Length	Relative Contact Order	MELD x MD Result
1aa3	63	0.10	0
1b22	70	0.11	1
1bax	94	0.12	1
1bdc	60	0.09	0
1d8j	81	0.09	0
1dp3	55	0.07	1
1ghh	81	0.12	0
1hdn	85	0.18	0
1i35	95	0.14	1
1iba	78	0.11	1
1iqo	88	0.07	1
1j27	98	0.19	0
1jw2	72	0.08	0
1k8b	52	0.15	0
1kjk	49	0.11	0
1kn6	73	0.21	0
1ln4	98	0.13	1
1lre	81	0.09	1
1oqk	78	0.12	1
1pul	103	0.08	1
1pv0	46	0.10	0
1qqv	67	0.09	1
1r1b	56	0.12	0
1r4g	53	0.07	0
1r5e	105	0.09	0
1rrz	66	0.09	1
1ryk	69	0.11	0
1sg7	75	0.11	0
1tkn	110	0.07	1
1w09	92	0.08	1
2ezh	65	0.10	1
2ezk	93	0.07	1
1gyz	60	0.10	0
1hyw	58	0.13	0
1kaf	108	0.10	0
1pc0	61	0.17	1
1rq6	62	0.08	0
1a6s	87	0.11	1
1eo0	77	0.09	1
1k5o	86	0.17	1
1nd9	49	0.16	1