Supporting Information

Clustering by RMSD and hydrogen-bonding pattern

To further explore the effects of clustering, clusters were defined using clustering criteria other than the ones discussed in the main text. In addition to the all-atom RMSD (see main text), the hydrogen bond pattern is used as an additional clustering criterion. The hydrogen-bond pattern is used because hydrogen bonds provide a relatively large favorable energy contribution and they reflect some of the finer details of the peptide packing. In the clustering process, a hydrogen bond is assumed to exist when the donor hydrogen and acceptor oxygen atoms are less than 2.4 Å apart.

The RMSD cutoff is set to 2 Å and the hydrogen bond pattern difference threshold is set to 1, i.e. the maximum hydrogen bond pattern difference is not allowed to be larger than 1. No angular restriction is applied for hydrogen bonds, only main-chain hydrogen bonds are taken into account, and hydrogen bonds between neighboring residues are excluded from the analysis. Using this clustering algorithm, the largest 30 clusters are selected to test the entropy estimators with the detailed data listed in Table S6. The results suggest that the relative performance of the entropy estimators remain unchanged with respect to the clustering that does not include the hydrogen-bonding criterion (described in the main text and in Tables S1-S4).

Table S1. TS (in units of kcal/mal) calculated from MD, quasi-harmonic approach using Cartesian coordinates, and quasi-harmonic approach using dihedral angles

Index	$S_{MD}T$	$S_{QH}T$	$S_{QH,dihe}T$
9	-8.26	-35.75	-5.45
8	-8.24	-41.54	-6.61
15	-8.04	-34.16	-3.42
10	-7.24	-19394	-4.87
12	-6.91	-22.75	-2.47
4	-6.80	-11.34	-3.21
16	-5.94	-28.55	-2.81
17	-5.37	-37.08	-3.85
5	-4.38	-3.67	-1.79
7	-4.27	-36.14	-2.40
14	-1.31	-17.21	1.27
11	-0.85	-14.33	1.46
19	-0.62	-8.71	0.31
6	-0.57	9.04	0.37
13	-0.54	-2.84	0.63
20	-0.47	-7.27	-0.20
3	-0.23	-4.53	1.60
2	-0.18	-3.31	-0.08
1	0	0	0
18	0.17	-3.31	0.99
E	16.44	110.556	22.54

Table S2. TS (kcal/mol) from MD and S_{1D}

Index	$S_{MD}T$	σ=20°	σ=10°	σ=5°	σ=2°	σ=1°	σ=0.5°
9	-8.26	-2.26	-3.31	-3.94	-4.41	-4.72	-5.20
8	-8.24	-2.24	-3.94	-5.10	-5.88	-6.21	-6.62
15	-8.04	-0.74	-1.07	-1.39	-1.84	-2.32	-3.14
10	-7.24	-1.79	-2.82	-3.42	-3.83	-4.16	-4.71
12	-6.91	-0.49	-0.87	-1.20	-1.59	-1.98	-2.61
4	-6.80	-0.74	-1.07	-1.30	-1.51	-1.69	-2.00
16	-5.94	-1.04	-1.31	-1.48	-1.84	-2.35	-3.27
17	-5.37	-1.14	-1.81	-2.30	-2.89	-3.52	-4.58
5	-4.38	-0.74	-0.96	-1.02	-1.08	-1.21	-1.49
7	-4.27	-1.08	-1.54	-1.89	-2.23	-2.50	-2.94
14	-1.31	2.23	2.87	3.09	2.91	2.47	1.64
11	-0.85	1.94	2.52	2.81	2.79	2.48	1.86
19	-0.62	1.20	1.76	1.96	1.71	1.11	-0.05
6	-0.57	1.16	1.56	1.74	1.70	1.55	1.23
13	-0.54	1.17	1.38	1.45	1.28	0.94	0.28
20	-0.47	0.80	1.11	1.12	0.72	0.066	-1.09
3	-0.23	1.47	1.86	2.17	2.31	2.24	2.03
2	-0.18	0.26	0.45	0.58	0.60	0.55	0.46
1	0	0	0	0	0	0	0
18	0.17	1.38	2.00	2.32	2.17	1.58	0.38
E	16.44	14.11	19.33	22.26	23.53	23.65	23.52

Table S3. TS (kcal/mol) from MD and S_{2D}

Index	$S_{MD}T$	σ=0.5	σ=0.2	σ=0.1	σ=0.05	σ=0.02
9	-8.26	-2.53	-5.78	-7.68	-9.43	-13.75
8	-8.24	-2.43	-6.10	-9.24	-12.35	-17.28
15	-8.04	-1.36	-2.16	-1.95	-2.64	-8.83
10	-7.24	-2.39	-5.97	-8.32	-10.33	-14.93
12	-6.91	-1.06	-1.06	0.04	0.16	-4.52
4	-6.80	-1.68	-3.86	-4.85	-5.52	-7.91
16	-5.94	-1.44	-2.44	-2.24	-2.99	-9.92
17	-5.37	-2.16	-3.81	-4.39	-5.86	-13.15
5	-4.38	-1.01	-1.56	-1.12	-0.90	-3.11
7	-4.27	-1.67	-3.58	-4.38	-5.22	-8.73
14	-1.31	0.94	2.98	4.79	4.99	-0.97
11	-0.85	0.40	1.91	3.64	4.16	-0.34
19	-0.62	0.43	2.22	3.94	3.59	-4.86
6	-0.57	0.13	1.61	3.47	4.48	2.37
13	-0.54	-0.16	0.67	2.00	2.19	-2.65
20	-0.47	-0.069	0.96	2.38	2.02	-6.22
3	-0.23	0.59	1.78	2.88	3.37	1.66
2	-0.18	-0.31	-0.23	0.29	0.78	0.31
1	0	0	0	0	0	0
18	0.17	-0.14	1.60	3.80	4.02	-3.81
E	16.44	9.13	22.96	32.48	37.21	37.07

Table S4. TS (kcal/mol) from MD and $S_{1D,nc}$

Index	$S_{MD}T$	$\sigma=20^o$	$\sigma=10^o$	$\sigma=5^o$	$\sigma=2^o$	$\sigma=1^o$	$\sigma=0.5^o$
9	-8.26	-3.08	-4.67	-5.61	-6.10	-6.41	-6.89
8	-8.24	-4.27	-6.84	-8.32	-8.99	-9.26	-9.61
15	-8.04	-2.63	-4.14	-4.98	-5.52	-5.96	-6.73
10	-7.24	-2.80	-4.36	-5.29	-5.77	-6.08	-6.58
12	-6.91	-1.86	-3.04	-3.62	-3.99	-4.32	-4.89
4	-6.80	-2.02	-3.16	-3.74	-4.02	-4.19	-4.47
16	-5.94	-2.20	-3.20	-3.76	-4.24	-4.78	-5.71
17	-5.37	-2.77	-4.04	-4.74	-5.29	-5.86	-6.88
5	-4.38	-1.21	-1.82	-2.20	-2.42	-2.57	-2.83
7	-4.27	-2.40	-3.76	-4.43	-4.76	-5.00	-5.39
14	-1.31	1.42	1.95	2.14	1.96	1.57	0.84
11	-0.85	0.064	-0.43	-0.68	-0.90	-1.18	-1.72
19	-0.62	0.48	0.26	0.048	-0.35	-0.95	-2.06
6	-0.57	0.20	0.25	0.24	0.18	0.045	-0.22
13	-0.54	-0.0014	-0.068	-0.12	-0.32	-0.64	-1.24
20	-0.47	-0.016	-0.026	-0.14	-0.53	-1.13	-2.20
3	-0.23	0.92	1.09	1.16	1.12	1.01	0.80
2	-0.18	-0.6	-0.36	-0.33	-0.34	-0.38	-0.47
1	0	0	0	0	0	0	0
18	0.17	0.80	0.82	0.72	0.29	-0.40	-1.58
E	16.44	12.43	14.08	14.62	14.71	14.63	14.44

Table S5. Correlation coefficient and slope of the fitted line for $S_{2D,NC}$ vs. S_{MD}

	Clusters 1-20		Clusters 1-20 and E	
σ[a]	Correlation	Slope	Correlation	Slope
0.5	0.913	0.289	0.966	0.372
0.2	0.930	0.601	0.974	0.618
0.1	0.933	0.733	0.971	0.694
0.05	0.933	0.781	0.970	0.732
0.02	0.930	0.794	0.969	0.747

a. σ in the $S_{2D,NC}$ method is dimensionless.

Table S6. Entropies using RMSD plus hydrogen bond pattern as similarity measure for clustering

Index	$S_{MD}T$	$S_{QH}T$	$S_{QH,dihe}T$	$S_{1D}T$
12	-5.29	-38.41	-6.18	-5.03
22	-4.30	-23.04	-4.09	-3.24
16	-4.16	-18.04	-4.11	-3.05
6	-3.92	-8.64	-3.29	-2.13
18	-3.83	-23.652	-2.42	-0.93
11	-3.54	-24.516	-3.91	-2.99
5	-3.14	-1.08	-2.53	-1.58
14	-2.20	-15.44	-2.12	-1.03
10	-1.51	-35.82	-3.81	-3.21
21	-1.18	-42.08	-4.80	-4.69
29	-1.03	-21.71	-2.70	-1.48
1	0.00	0.00	0.00	0.00
2	0.90	-3.74	-0.27	0.88
4	1.80	5.22	1.52	2.45
3	1.91	-1.08	1.65	3.14
9	2.38	-10.66	1.56	3.30
24	2.44	-6.48	-0.50	1.46
7	2.62	1.01	-0.14	0.52
17	2.70	-3.42	0.41	1.79
28	2.88	-17.82	-0.62	0.61
19	3.17	-5.76	0.084	2.09
20	3.54	-5.87	0.29	2.47
8	3.57	-0.68	0.116	2.42
13	3.81	-0.25	3.13	4.86
27	4.05	-14.29	1.12	3.18
15	4.10	16.88	0.022	2.16
25	4.22	-7.74	0.95	2.75
23	4.27	5.62	0.69	2.28
26	4.71	0.40	2.43	3.49
30	5.20	-6.66	1.31	3.12

Column 1 is the cluster index. Columns 2-5 show the various entropies multiplied by T. (in units of kcal/mol.) Column 2 corresponds to the MD simulation and clustering. Column 3 corresponds to quasi-harmonic approach using Cartesian coordinates. Column 4 corresponds to quasi-harmonic approach using dihedral angles. Column 5 corresponds to S_{1D} with smoothing parameter $\sigma = 5^{o}$.