

Supplementary Information

for

UniCon3D: de novo protein structure prediction using united-residue conformational search via stepwise, probabilistic sampling

Debswapna Bhattacharya¹, Renzhi Cao¹ and Jianlin Cheng^{1, 2, 3, *}

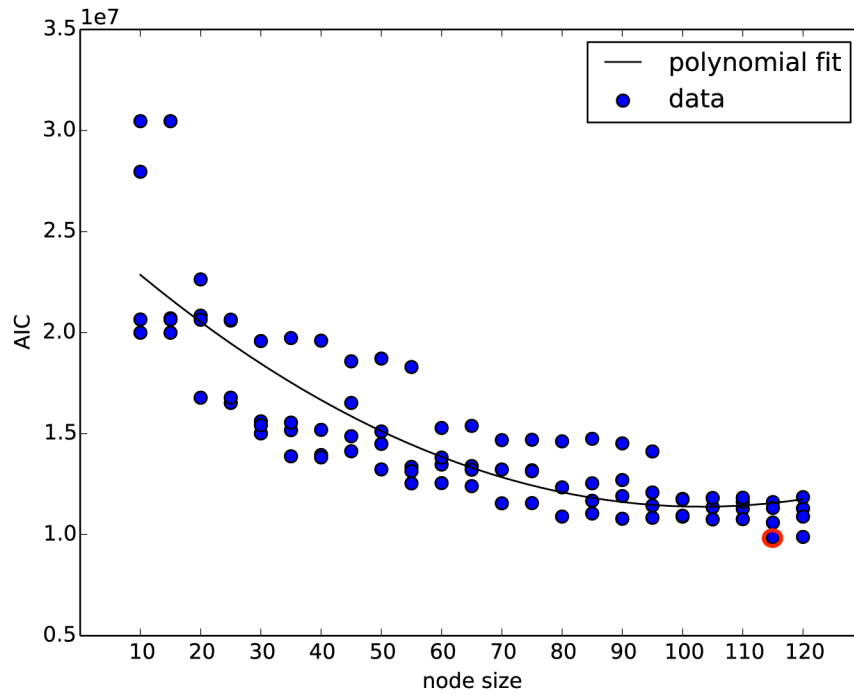
¹Department of Computer Science, University of Missouri, Columbia, MO 65211, USA

²Informatics Institute, University of Missouri, Columbia, MO 65211, USA

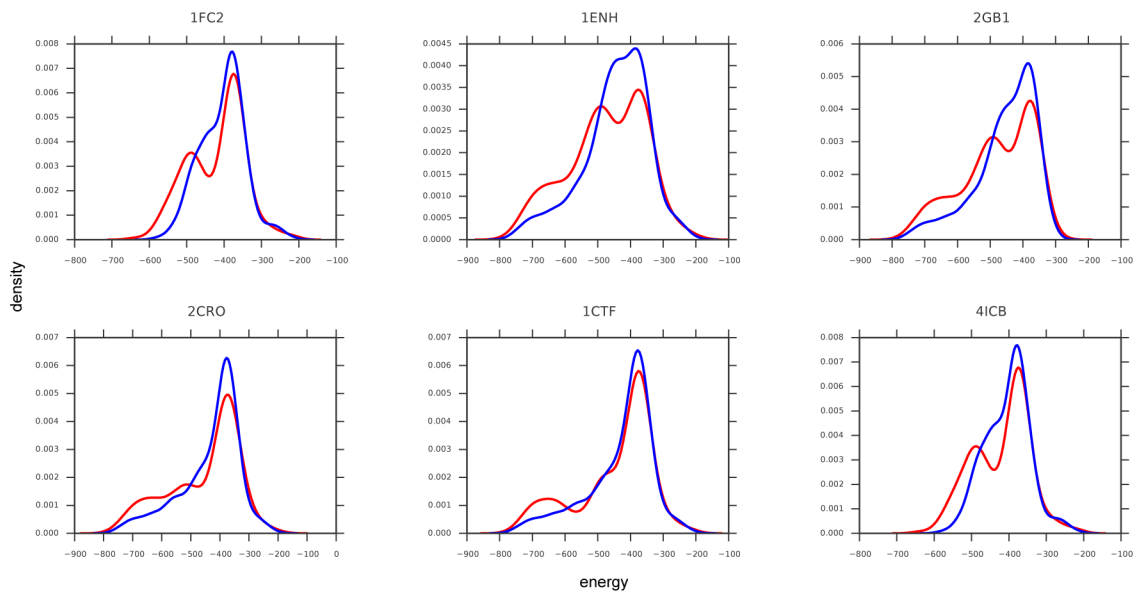
³C. Bond Life Science Center, University of Missouri, Columbia, MO 65211, USA

*To whom correspondence should be addressed. Phone: (573)-882-7306. Fax: (573)-882-8318. E-mail: chengji@missouri.edu.

Supplementary Item	Title
Supplementary Figure 1	Optimal model selection using AIC.
Supplementary Figure 2	Density of energy sampled by UniCon3D and B ₀ 3D.
Supplementary Table 1	TM-scores and C α -rmsds for UniCon3D and top-performing servers in 30 CASP11 FM domains.
Supplementary Table 2	TM-scores and C α -rmsds of top-ranked decoys based on different MQAPs for each of the 30 CASP11 FM domains.
Supplementary Table 3	TM-scores and C α -rmsds for UniCon3D and top-performing servers in 15 CASP10 FM domains.
Supplementary Table 4	Target-by-target comparison of TM-scores and C α -rmsds between UniCon3D, B03D, GDFuzz3D and FT-COMAR in 45 CASP10 TBM targets. Best CASP10 results are also provided as a reference.
Supplementary Table 5	Target-by-target comparison of TM-scores and C α -rmsds between UniCon3D, GDFuzz3D and FT-COMAR in 10 CASP10 TBM targets for which MULTICOM produced contact map with sensitivity > 0.5. Best CASP10 results are also provided as a reference.



Supplementary Figure 1. Optimal model selection using AIC. AIC values versus varying hidden node sizes are shown, with four models trained for each hidden node size. The curved line is tendency line constructed by fitting second degree polynomial to the data. The minimum AIC value corresponds to the optimal model (highlighted in red circle).



Supplementary Figure 2. Density of energy sampled by UniCon3D and B₀3D. Kernel density estimates of sampled energy for UniCon3D (red) and B₀3D (blue) considering 100 decoys in a benchmark of six small proteins. PDB code for each protein target is shown at the top.

Supplementary Table 1. TM-scores and C α -rmsds for UniCon3D and top-performing servers in 30 CASP11 FM domains.

Target	QUARK		Zhang-Server		UniCon3D		RBO_Aleph		nns		BAKER-ROSETTASERVER	
	TM-score	C α -rmsd	TM-score	C α -rmsd	TM-score	C α -rmsd	TM-score	C α -rmsd	TM-score	C α -rmsd	TM-score	C α -rmsd
T0761-D1	0.18	14.60	0.18	15.02	0.23	11.69	0.256	14.748	0.24	12.58	0.24	14.51
T0761-D2	0.28	16.42	0.28	16.22	0.19	15.82	0.205	18.292	0.24	19.74	0.23	18.45
T0763-D1	0.19	16.03	0.20	15.84	0.20	19.15	0.211	19.798	0.17	18.01	0.20	16.64
T0767-D2	0.26	17.28	0.33	14.23	0.21	21.08	0.200	17.651	0.23	18.65	0.26	16.76
T0771-D1	0.25	13.25	0.25	13.03	0.21	19.13	0.315	13.693	0.19	18.64	0.19	17.47
T0777-D1	0.31	14.77	0.24	20.35	0.23	21.60	0.265	18.998	0.21	23.52	0.21	22.78
T0781-D1	0.21	19.41	0.18	26.03	0.19	18.12	0.251	17.265	0.18	20.55	0.19	20.29
T0785-D1	0.23	15.79	0.26	13.79	0.23	14.91	0.209	15.421	0.21	17.44	0.21	14.27
T0789-D1	0.31	15.92	0.28	15.50	0.29	12.56	0.183	21.293	0.22	16.84	0.22	20.87
T0789-D2	0.20	14.71	0.21	13.67	0.24	14.61	0.175	15.668	0.19	18.31	0.20	16.24
T0790-D1	0.32	15.00	0.28	15.70	0.34	12.20	-	-	0.23	15.12	0.22	19.95
T0790-D2	0.20	14.48	0.20	14.06	0.22	14.56	-	-	0.21	17.38	0.20	18.17
T0791-D1	0.28	13.29	0.26	17.49	0.29	13.40	0.279	15.058	0.19	15.31	0.18	18.45
T0791-D2	0.31	14.21	0.32	14.38	0.29	11.71	0.292	12.831	0.20	17.36	0.19	14.89
T0794-D2	0.26	12.88	0.30	12.49	0.18	18.65	0.279	20.948	0.18	20.00	0.25	17.63
T0806-D1	0.25	17.48	0.26	16.34	0.25	16.89	0.260	17.951	0.21	17.89	0.19	21.18
T0808-D2	0.22	19.23	0.23	19.20	0.21	22.55	0.194	21.693	0.27	21.01	0.21	23.40
T0810-D1	0.32	13.14	0.42	8.30	0.20	15.07	0.176	13.024	0.21	19.50	0.29	14.76
T0814-D1	0.25	18.24	0.26	16.70	0.19	16.78	0.181	20.732	0.42	10.47	0.24	20.88
T0814-D2	0.21	21.09	0.37	12.56	0.32	18.25	0.193	17.419	0.22	22.28	0.29	24.48
T0820-D1	0.36	8.83	0.38	9.32	0.23	12.39	0.301	14.926	0.33	11.65	0.27	14.75
T0824-D1	0.30	14.47	0.30	14.75	0.30	13.11	0.251	12.796	0.26	14.65	0.27	14.01
T0827-D2	0.38	8.89	0.19	18.79	0.24	15.77	0.230	14.629	0.28	15.54	0.23	15.11
T0831-D2	0.27	20.39	0.40	15.74	0.19	20.70	0.272	19.433	0.23	25.97	0.21	25.11
T0832-D1	0.22	18.80	0.22	19.72	0.21	19.74	0.232	22.041	0.21	19.46	0.18	19.68
T0834-D1	0.27	17.21	0.20	20.22	0.27	17.20	0.347	16.426	0.17	22.95	0.20	19.32
T0834-D2	0.34	11.17	0.28	12.58	0.29	10.51	0.251	14.768	0.29	10.57	0.22	11.19
T0836-D1	0.27	15.63	0.28	22.47	0.27	14.71	0.269	17.548	0.28	12.26	0.25	19.55
T0837-D1	0.74	2.95	0.43	9.27	0.35	10.68	0.441	11.311	0.30	14.77	0.27	11.90
T0855-D1	0.42	9.38	0.48	7.85	0.32	10.93	0.421	10.507	0.27	13.92	0.43	8.51
Mean	0.29	14.83	0.28	15.39	0.25	15.82	0.25	16.67	0.23	17.41	0.23	17.71

Supplementary Table 2. TM-scores and C α -rmsds of top-ranked decoys based on different MQAPs for each of the 30 CASP11 FM domains.

Target	UniCon3D		APOLLO		Qprob		ProQ2		MUFOLD-CL	
	TM-score	C α -rmsd	TM-score	C α -rmsd	TM-score	C α -rmsd	TM-score	C α -rmsd	TM-score	C α -rmsd
T0761-D1	0.23	11.69	0.24	12.21	0.19	14.55	0.22	13.05	0.24	13.555
T0761-D2	0.19	15.82	0.20	15.04	0.24	14.32	0.22	14.25	0.25	15.013
T0763-D1	0.20	19.15	0.26	15.50	0.19	18.02	0.19	18.02	0.22	16.371
T0767-D2	0.21	21.08	0.20	17.85	0.19	22.00	0.19	22.00	0.21	19.170
T0771-D1	0.21	19.13	0.19	17.53	0.21	19.13	0.19	20.62	0.20	18.201
T0777-D1	0.23	21.60	0.23	21.60	0.28	16.95	0.20	22.52	0.23	22.879
T0781-D1	0.19	18.12	0.27	18.74	0.23	17.54	0.16	20.47	0.18	20.657
T0785-D1	0.23	14.91	0.24	15.43	0.22	13.38	0.22	12.88	0.20	12.192
T0789-D1	0.29	12.56	0.27	13.15	0.34	13.49	0.34	15.23	0.26	12.656
T0789-D2	0.24	14.61	0.21	16.46	0.27	12.25	0.20	15.16	0.24	16.232
T0790-D1	0.34	12.20	0.30	11.68	0.24	15.98	0.26	12.25	0.30	13.971
T0790-D2	0.22	14.56	0.24	15.68	0.26	16.53	0.28	16.77	0.24	15.927
T0791-D1	0.29	13.40	0.28	11.85	0.26	14.01	0.27	12.55	0.23	13.981
T0791-D2	0.29	11.71	0.23	12.35	0.22	15.56	0.26	17.74	0.27	12.756
T0794-D2	0.18	18.65	0.18	18.65	0.21	20.21	0.19	17.92	0.20	17.054
T0806-D1	0.25	16.89	0.28	17.23	0.21	20.25	0.21	20.25	0.25	16.891
T0808-D2	0.21	22.55	0.21	21.61	0.22	21.91	0.21	20.11	0.20	23.675
T0810-D1	0.20	15.07	0.17	16.05	0.23	14.34	0.19	14.63	0.18	14.546
T0814-D1	0.19	16.78	0.18	15.90	0.20	18.96	0.18	15.90	0.18	19.161
T0814-D2	0.32	18.25	0.20	17.88	0.19	22.39	0.20	17.88	0.20	21.940
T0820-D1	0.23	12.39	0.31	10.29	0.33	10.82	0.28	13.28	0.27	15.130
T0824-D1	0.30	13.11	0.26	10.85	0.24	12.12	0.21	14.29	0.24	12.887
T0827-D2	0.24	15.77	0.23	17.28	0.25	14.45	0.32	14.44	0.26	15.730
T0831-D2	0.19	20.70	0.24	18.97	0.19	26.43	0.27	16.53	0.22	22.177
T0832-D1	0.21	19.74	0.24	21.95	0.25	17.56	0.24	18.39	0.24	20.966
T0834-D1	0.27	17.20	0.21	21.11	0.23	12.48	0.21	19.09	0.21	21.532
T0834-D2	0.29	10.51	0.22	12.71	0.27	10.36	0.23	12.51	0.24	13.462
T0836-D1	0.27	14.71	0.29	17.00	0.21	19.03	0.21	17.46	0.26	17.356
T0837-D1	0.35	10.68	0.35	11.25	0.32	11.67	0.23	12.61	0.28	10.750
T0855-D1	0.32	10.93	0.28	12.61	0.27	13.57	0.30	9.72	0.26	12.184
Mean	0.25	15.82	0.24	15.88	0.24	16.34	0.23	16.28	0.23	16.63

Supplementary Table 3. TM-scores and Ca-rmsds for UniCon3D and top-performing servers in 15 CASP10 FM domains.

Target	Zhang-Server		QUARK		UniCon3D		PMS		MUFold_CRF		BAKER-ROSETTASERVER	
	TM-score	Ca-rmsd	TM-score	Ca-rmsd	TM-score	Ca-rmsd	TM-score	Ca-rmsd	TM-score	Ca-rmsd	TM-score	Ca-rmsd
T0658-D1	0.17	19.60	0.18	19.63	0.22	19.23	0.20	18.71	0.18	18.61	0.17	21.95
T0666-D1	0.40	7.22	0.35	8.75	0.24	15.17	0.27	15.20	0.24	15.57	0.26	17.62
T0684-D2	0.23	17.44	0.22	17.68	0.22	20.28	0.21	17.51	0.24	16.54	0.22	18.37
T0693-D1	0.31	14.24	0.31	14.22	0.22	13.19	0.25	11.40	-	-	0.21	13.88
T0695-D1	0.17	39.60	0.17	39.46	0.16	46.12	0.13	44.13	0.11	51.02	0.17	64.73
T0719-D6	0.21	17.81	0.19	17.31	0.20	14.86	0.16	22.68	0.32	16.60	0.20	20.35
T0726-D3	0.21	14.56	0.17	14.87	0.17	14.83	0.19	12.63	0.19	14.32	0.15	17.89
T0734-D1	0.24	17.62	0.17	22.04	0.18	20.78	0.20	20.30	0.24	17.10	0.17	21.22
T0735-D2	0.34	12.90	0.35	12.93	0.24	18.38	0.25	13.25	0.25	12.52	0.31	15.36
T0737-D1	0.32	10.92	0.32	10.81	0.27	12.13	0.29	13.05	0.26	15.14	0.21	12.76
T0739-D1	0.25	10.59	0.31	8.90	0.29	9.21	0.22	13.34	0.18	13.97	0.19	14.89
T0739-D2	0.22	7.73	0.22	7.79	0.23	9.75	0.17	10.99	0.19	12.99	0.17	10.05
T0740-D1	0.27	20.64	0.26	20.42	0.23	16.76	0.34	16.85	0.15	20.26	0.21	53.63
T0741-D1	0.14	18.11	0.15	18.33	0.15	19.98	0.13	20.21	0.14	18.89	0.13	19.37
T0756-D2	0.33	13.13	0.32	13.14	0.28	11.12	0.29	20.32	0.26	18.25	0.35	9.27
Mean	0.25	16.14	0.25	16.42	0.22	17.45	0.22	18.04	0.21	18.70	0.21	22.09

Supplementary Table 4. Target-by-target comparison of TM-scores and Ca-rmsds between UniCon3D, B₀3D, GDFuzz3D and FT-COMAR in 45 CASP10 TBM targets. Best CASP10 results are also provided as a reference.

	UniCon3D		B ₀ 3D		GDFuzz3D		FT-COMAR		Best CASP10	
	TM-score	Ca-rmsd	TM-score	Ca-rmsd	TM-score	Ca-rmsd	TM-score	Ca-rmsd	TM-score	Ca-rmsd
T0644	0.43	9.88	0.43	11.97	0.63	4.70	0.51	7.50	0.88	2.22
T0649	0.32	20.04	0.35	13.26	0.23	15.50	0.38	13.60	0.46	11.98
T0661	0.47	10.42	0.43	12.36	0.43	8.00	0.37	10.00	0.89	2.74
T0662	0.62	3.19	0.60	6.37	0.55	3.40	0.52	3.60	0.88	1.54
T0667	0.59	7.35	0.46	8.40	0.74	6.10	0.29	14.90	0.84	3.94
T0669	0.49	10.05	0.54	5.18	0.58	5.10	0.43	7.50	0.73	2.91
T0676	0.37	16.10	0.34	21.59	0.19	16.10	0.16	24.90	0.51	7.61
T0678	0.46	7.79	0.51	6.89	0.39	8.20	0.24	13.60	0.46	6.67
T0688	0.55	6.48	0.47	8.00	0.43	7.70	0.30	10.90	0.88	2.52
T0703	0.40	12.15	0.37	16.58	0.19	3.40	0.18	4.90	0.90	2.48
T0704	0.42	18.00	0.33	16.06	0.40	10.90	0.20	15.80	0.87	2.38
T0720	0.38	16.78	0.37	11.15	0.08	7.70	0.07	15.30	0.68	4.66
T0721	0.43	13.13	0.28	16.61	0.18	12.50	0.13	15.90	0.94	2.04
T0746	0.32	17.55	0.28	18.44	0.32	25.70	0.21	26.20	0.70	7.42
T0747	0.39	11.49	0.38	14.07	0.32	14.50	0.29	16.40	0.73	12.15
T0752	0.50	12.09	0.46	11.68	0.76	4.80	0.59	6.60	0.93	1.50
T0753	0.54	4.68	0.47	12.15	0.18	4.10	0.17	5.00	0.84	2.04
T0650	0.52	8.37	0.37	12.44	0.36	8.40	0.20	17.00	0.98	1.31
T0654	0.44	7.52	0.34	17.08	0.78	4.20	0.69	3.50	0.83	3.21
T0655	0.52	6.90	0.40	9.49	0.62	7.70	0.51	32.50	0.84	3.97

T0657	0.50	6.48	0.50	6.61	0.70	5.00	0.61	7.60	0.90	2.25
T0659	0.54	3.42	0.50	3.84	0.38	9.50	0.18	14.30	0.94	0.86
T0672	0.36	15.02	0.33	17.12	0.77	6.60	0.29	18.20	0.81	5.19
T0673	0.31	11.74	0.30	11.85	0.22	10.90	0.22	10.20	0.55	5.35
T0679	0.53	7.26	0.46	9.58	0.77	8.30	0.65	8.30	0.82	3.84
T0680	0.56	4.81	0.61	3.85	0.16	18.40	0.11	24.00	0.73	3.39
T0681	0.36	12.49	0.35	16.64	0.06	19.70	0.05	18.20	0.88	2.14
T0682	0.56	6.62	0.54	8.11	0.22	19.60	0.20	19.60	0.93	1.62
T0689	0.41	15.44	0.38	11.51	0.16	13.20	0.08	19.60	0.94	1.53
T0691	0.35	16.23	0.34	11.66	0.08	17.70	0.08	20.10	0.50	7.55
T0699	0.44	13.61	0.34	16.33	0.42	7.30	0.20	16.30	0.93	1.95
T0700	0.55	7.72	0.54	7.34	0.12	16.00	0.11	20.20	0.83	0.80
T0707	0.46	14.76	0.38	15.75	0.17	24.90	0.11	29.60	0.74	7.39
T0708	0.41	11.18	0.39	10.72	0.42	8.60	0.39	7.10	0.90	3.17
T0712	0.38	14.19	0.32	14.35	0.43	14.20	0.39	13.00	0.94	1.90
T0714	0.43	4.92	0.47	6.25	0.70	2.70	0.54	3.80	0.91	1.14
T0716	0.70	2.13	0.70	2.25	0.55	1.00	0.41	8.00	0.92	0.73
T0731	0.68	2.19	0.68	2.22	0.48	8.40	0.20	11.70	0.82	1.81
T0733	0.40	15.12	0.30	21.06	0.15	27.20	0.15	27.20	0.91	3.10
T0738	0.52	8.82	0.50	10.57	0.88	2.40	0.72	4.30	0.96	1.40
T0742	0.29	16.46	0.32	15.83	0.18	27.20	0.13	35.70	0.70	5.35
T0743	0.44	9.01	0.46	9.41	0.32	12.90	0.29	22.90	0.76	3.85
T0744	0.32	19.93	0.29	22.32	0.51	10.60	0.41	20.70	0.75	5.78
T0755	0.39	11.85	0.36	12.10	0.62	9.70	0.31	15.00	0.92	3.21
T0757	0.41	11.35	0.35	16.59	0.83	2.90	0.64	5.30	0.93	1.89
Mean	0.45	10.73	0.42	11.86	0.41	10.75	0.31	14.81	0.82	3.61

Supplementary Table 5. Target-by-target comparison of TM-scores and C α -rmsds between UniCon3D, GDFuzz3D and FT-COMAR in 10 CASP10 TBM targets for which MULTICOM produced contact map with sensitivity > 0.5. Best CASP10 results are also provided as a reference.

Target	UniCon3D		B ₀ 3D		GDFuzz3D		FT-COMAR		Best CASP10	
	TM-score	C α -rmsd	TM-score	C α -rmsd	TM-score	C α -rmsd	TM-score	C α -rmsd	TM-score	C α -rmsd
T0662	0.62	3.19	0.60	6.37	0.55	3.40	0.52	3.60	0.88	1.54
T0669	0.49	10.05	0.54	5.18	0.58	5.10	0.43	7.50	0.73	2.91
T0688	0.55	6.48	0.47	8.00	0.43	7.70	0.30	10.90	0.88	2.52
T0753	0.54	4.68	0.47	12.15	0.18	4.10	0.17	5.00	0.84	2.04
T0650	0.52	8.37	0.37	12.44	0.36	8.40	0.20	17.00	0.98	1.31
T0655	0.52	6.90	0.40	9.49	0.62	7.70	0.51	32.50	0.84	3.97
T0657	0.50	6.48	0.50	6.61	0.70	5.00	0.61	7.60	0.90	2.25
T0659	0.54	3.42	0.50	3.84	0.38	9.50	0.18	14.30	0.94	0.86
T0714	0.43	4.92	0.47	6.25	0.70	2.70	0.54	3.80	0.91	1.14
T0716	0.70	2.13	0.70	2.25	0.55	7.90	0.41	8.00	0.92	0.73
Mean	0.54	5.66	0.50	7.26	0.51	6.15	0.39	11.02	0.88	1.93