Experimental Inferential Structure Determination of Ensembles for Intrinsically Disordered Proteins

David H. Brookes[2] and Teresa Head-Gordon[1,2,3,4*]

[1]Department of Chemistry, [2]Department of Bioengineering, [3]Department of Chemical and Biomolecular Engineering, [4]Chemical Sciences Division, Lawrence Berkeley National Laboratory, University of California, Berkeley, CA 94720.

Table S1: *Parameters of Gaussian distributions used to model each experimental nuisance parameter.* The Karplus equation coefficients, and their mean and variances, are taken directly from Vuister and Bax[1]. For chemical shifts we utilize SHIFTX2[2] as the back-calculator, which has well-documented error distributions.

Type	Nuisance Parameter	Mean	Variance
Karplus equation coefficients $\xi_{(back)}$ $J = A\cos^2\phi + B\cos\phi + C$	A	6.51	0.14
	B	-1.76	0.03
	C	1.60	0.08
SHIFTX2 $\xi_{(back)}$	HZ	0	0.068121
	C	0	0.25969216
	CE1	0	0.84052224
	CB	0	0.28398241
	CA	0	0.14714896
	HB2	0	0.03290596
	CG	0	0.44023225
	CE	0	0.136161
	N	0	1.51979584
	CZ	0	1.36258929
	CD1	0	1.02272769
	CD2	0	1.39523344
	CG1	0	0.61387225
	HG3	0	0.02852721
	HA	0	0.01168561
	HG	0	0.14638276
	HE	0	0.11168964
	HG2	0	0.01962801
	CG2	0	0.84309124
	HD3	0	0.03225616
	HD2	0	0.03297856
	HD1	0	0.030976
	HE2	0	0.025281
	HE3	0	0.10361961
	HB	0	0.02143296
	HG12	0	0.03538161
	HG1	0	0.02217121
	CD	0	0.25593481
	HG13	0	0.01610361
	H	0	0.05527201
	HE1	0	0.03798601
	HA2	0	0.07463824
	HA3	0	0.09042049
	HB3	0	0.04447881
	CE2	0	0.65480464

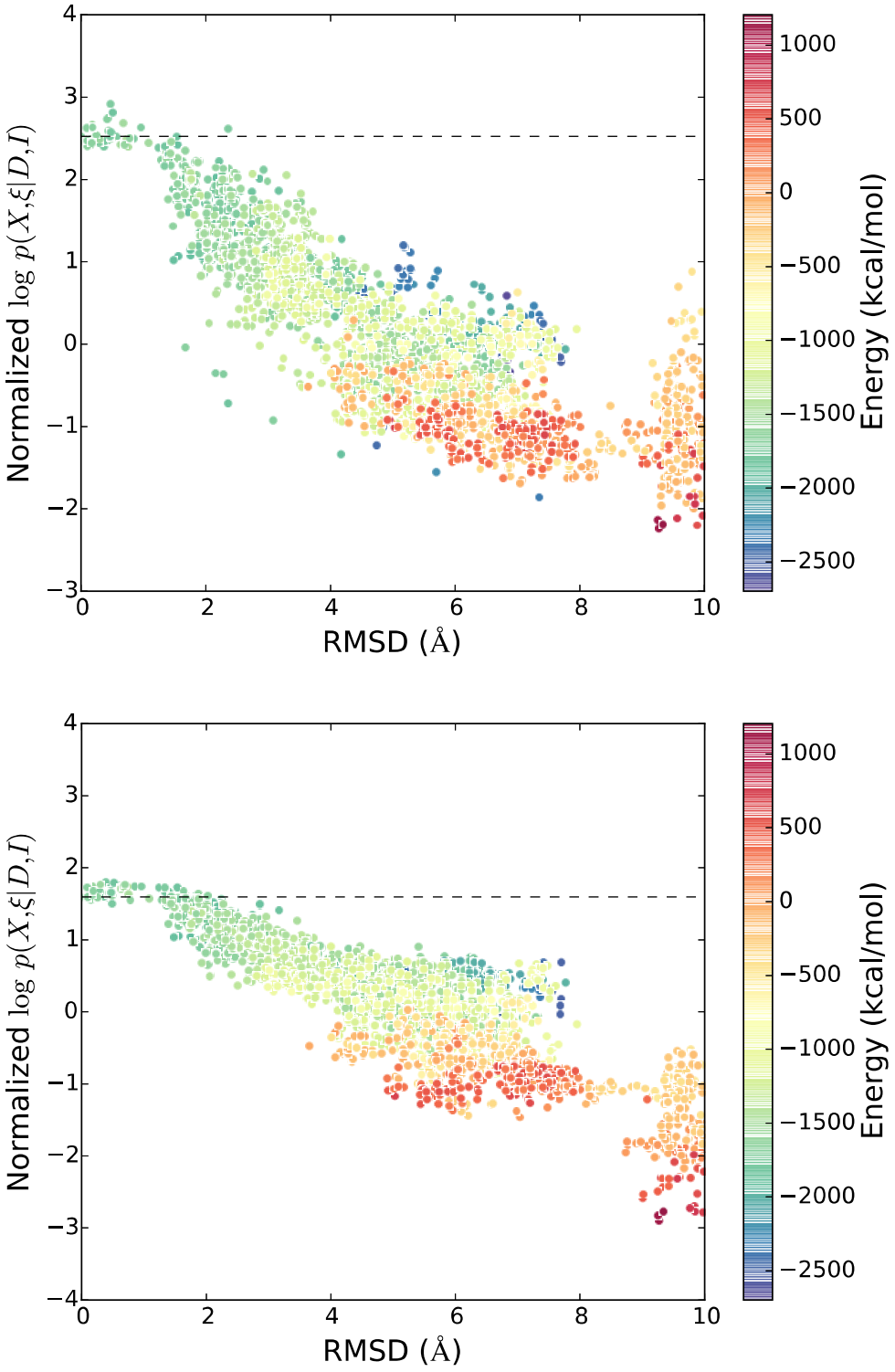

Figure S1: *$log\, p(X,\xi|D,I)$ vs. RMSD for ~25,000 misfolded structures for Trp-cage (PDB ID: 1L2Y) for (a) OISD and (b) EISD, both using a Boltzmann prior.* Dotted black lines represent the fit-to-data probability of the native structure. All probabilities were normalized so the set had a mean of 0 and a variance of 1 (for easier comparison between schemes).

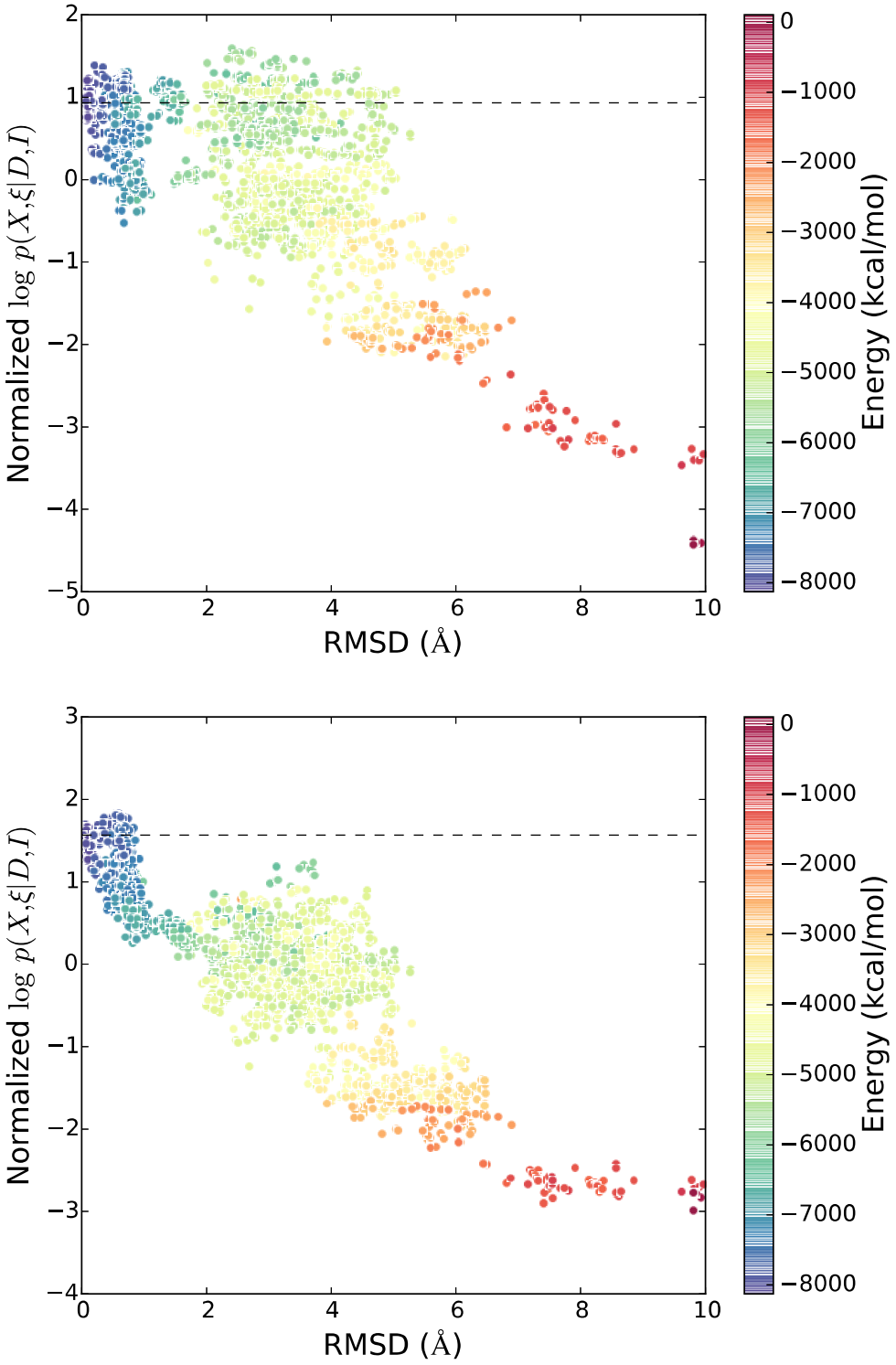

Figure S2: *$log\, p(X,\xi|D,I)$ vs. RMSD for ~25,000 misfolded structures for 1JH8 for (a) OISD and (b) EISD, both using a uniform prior.* Dotted black lines represent the fit-to-data probability of the native structure. All probabilities were normalized so the set had a mean of 0 and a variance of 1 (for easier comparison between schemes).

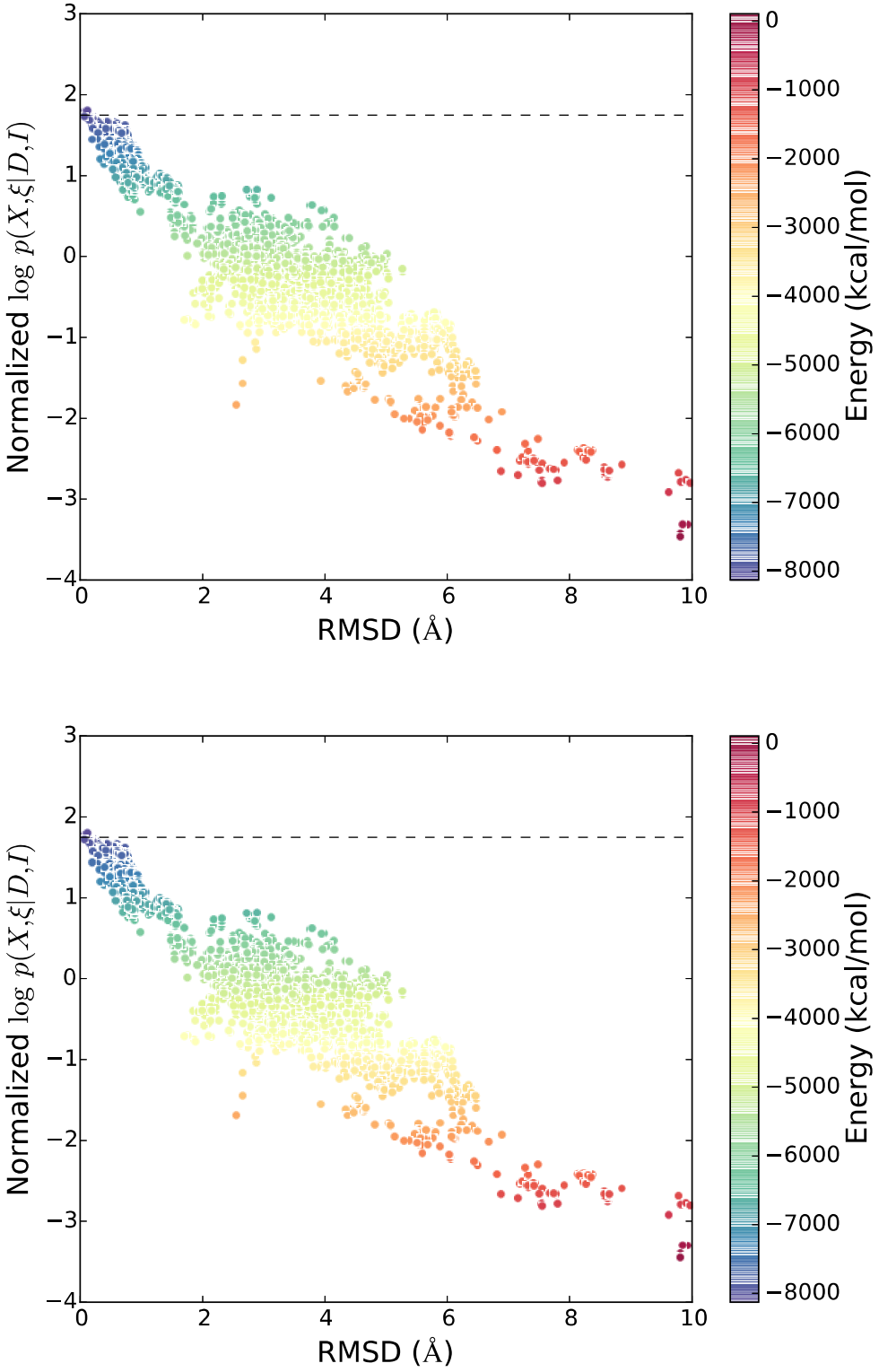

Figure S3: *$\log p(X,\xi|D,I)$ vs. RMSD for ~25,000 misfolded structures for 1JH8 for (a) OISD and (b) EISD, both using a Boltzmann prior.* Dotted black lines represent the fit-to-data probability of the native structure. All probabilities were normalized so the set had a mean of 0 and a variance of 1 (for easier comparison between schemes).

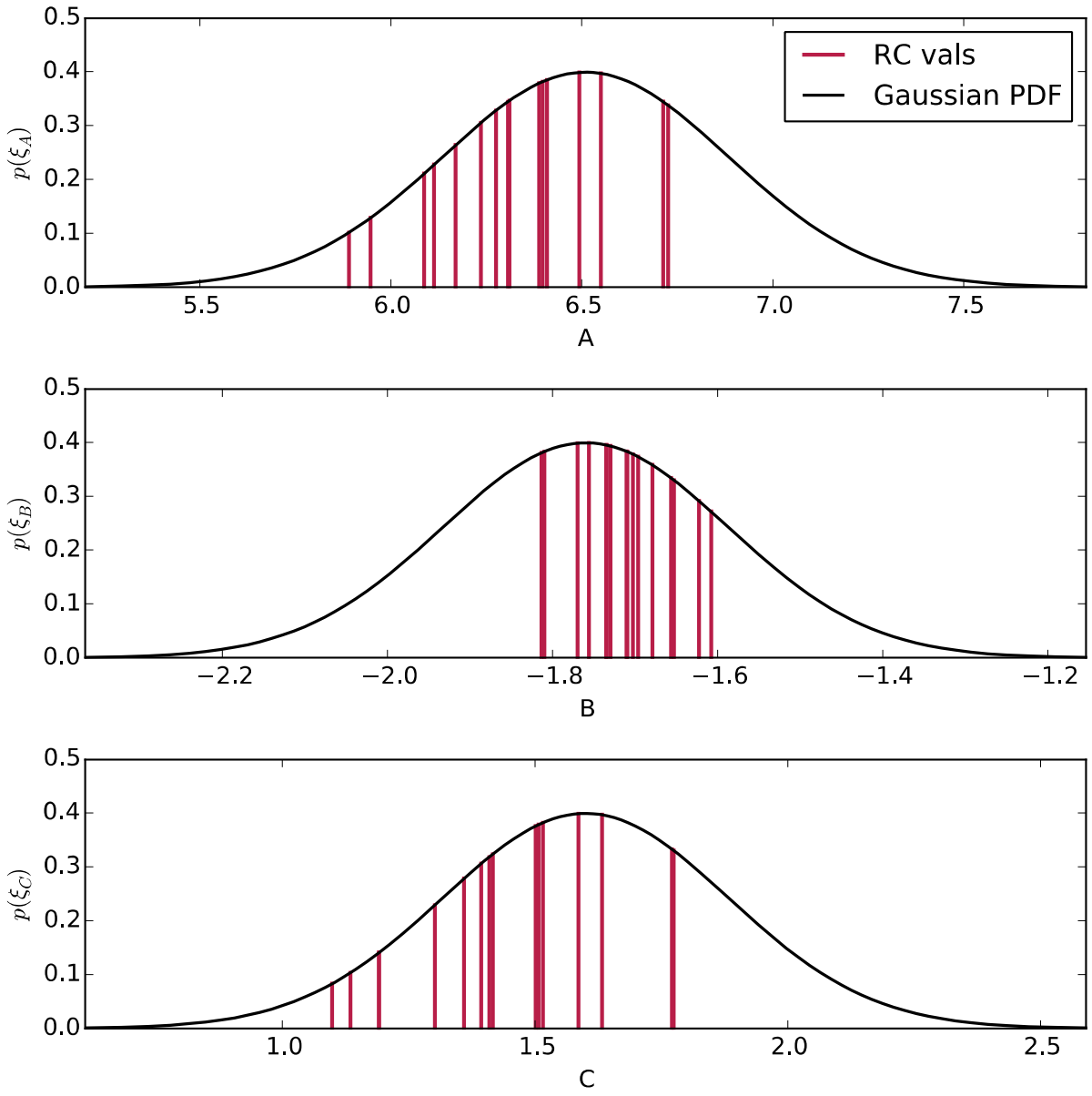

Figure S4: *Optimized coefficients of the Karplus equation that produce the highest EISD probability for a random coil ensemble (RC) for the amyloid-beta 42 (Aβ42) monomer.* The Gaussian PDF models the experimental Vuister and Bax scalar coupling constants A, B, and C used in the Karplus equation given in Table S1. The lines correspond to the optimized values of these constants used in the back-calculation from the RC structural ensemble to maximize fit-to-data probabilities using 16 $J_H{}^N{}_H{}^\alpha$ experimental measurements for Aβ42. Similar results are obtained for all other parameters for all Aβ42 ensembles.

REFERENCES

(1) Vuister, G. W.; Delaglio, F.; Bax, A. *J. Biomol. NMR* **1993**, *3*, 67.
(2) Han, B.; Liu, Y. F.; Ginzinger, S. W.; Wishart, D. S. *J. Biomol. NMR* **2011**, *50*, 43.