Supplemental Information: Realistic ensemble models of intrinsically disordered proteins using a structure-encoding coil database

Alejandro Estaña^{a,b}, Nathalie Sibille^b, Elise Delaforge^b, Marc Vaisset^a, Juan Cortés^a, Pau Bernadó^b

^aLAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France ^bCentre de Biochimie Structurale. INSERM, CNRS, Université de Montpellier, France

	SRS		TRS		Hybrid		
Protein	Q-factor	χ^2	Q-factor	χ^2	Q-factor	χ^2	07 Structure
	(RDCs)	(SAXS)	(RDCs)	(SAXS)	(RDCs)	(SAXS)	70 Structure
MKK7	100.36		67.98		45.20		29.00
eba181	56.01		114.33		46.90		8.18
p15	79.61	1.03	88.11	1.01	63.12	1.04	30.90
sic1	50.67		92.16		53.29		9.78
ntailMV	98.98		91.07		47.23		8.33
ntailSV	110.61		68.97		43.97		15.85
src	68.88	2.70	114.16	2.58	60.57	1.93	5.26
K18	63.15		83.61		59.32		8.34
Tau	63.32	2.02	77.87	2.15	60.37	1.52	18.55

Table S1: Related to Figures 1 and 2. Q-factor and χ^2 values obtained from the comparison of the experimental and computational data for each of the proteins studied.

 $Preprint\ submitted\ to\ Structure$

IDPs	TRS regions
MKK7	[A2-D29]
eba181	[D985-P986], [P997], [D1004-A1005], [D1015-D1016], [K1086-G1092]
p15	[Y13-S29], [W61-Q62], [R70-D75], [A94-H105]
sic1	[S26-L29], [P72-T77]
ntailMV	[R489-Q499]
ntailSV	[V476-E489]
src	[F64-S69]
K18	[L253-V256], [L284-N286], [L315-K317], [F346-K347]
Tau	[E115-H121], [R170-S184], [T212-P236], [L253-V256], [L284-N286], [L315-K317], [F346-K347]

Table S2: Related to Figure 2. Regions defined as partially structured in the hybrid strategy for the studied proteins.



Figure S1: Related to Figure 2. Experimental scattering intensity (black dots) in logarithmic scale as a function of the momentum transfer, s, compared with the averaged profiles computed from the hybrid ensemble models for tau (green), src (pink), and p15 (red). The profiles have been displaced along the y-axis for a better inspection.



Figure S2: Related to Figure 4. Conformational sampling for the four turns identified in K18. Each column displays the Ramachandran plots for the first three residues belonging to turns 1 to 4 when using the SRS (blue) or the hybrid (green) sampling approaches.



Figure S 3: Related to Figure 5. Deviation with respect to a canonical helix of conformations of the N-terminal region of MKK7 generated using different sampling strategies. The abscissa axis corresponds to the distance between the first N and last C backbone atoms of the 28-residue long structured region of MKK7. The ordinate axis corresponds to the average distance between H-bond donor and acceptor atoms. For a canonical 28-residue long helix these values are 3 Å and 43.5 Å, respectively. The proposed hybrid sampling strategy (left image), using TRS for the selected 28-residue region, produces a wide variety of conformations, and several regions of the plot are more densely sampled. In particular, we can observe more conformations near the coordinates corresponding to the canonical helix, but slightly shorter (in terms of end-to-end distance). This can be explained by the propensity to form shorter helices, as observed in the the SS-map, giving the possibility to flanking regions to fold back. Using the procedure that imposes helicity at the residue level (middle image), the distribution of the sampled conformations is much more compact, and the region near the canonical helix is not sampled at all. This is in line with the SS-map representation, showing that only very short helices are formed. Finally, the plot corresponding to the other procedure clearly shows the imposed 25% populated canonical helix (right image), being the rest of the conformations far from it. Such a discontinuity in the conformational space is unrealistic. Indeed, although this last procedure can in some cases yield a good agreement between computationally generated ensembles and experimental data, these ensembles are inaccurate representations of the conformational heterogeneity of partially structured regions in IDPs.



Figure S4: Related to Figure 2. Comparison of the experimental (black) $C\alpha$, $C\beta$, CO and NH chemical shifts of ntailMV with these obtained from the ensembles built with the SRS (blue) and hybrid (red) strategies. Orange bars indicate these regions defined as partially structured in the hybrid modeling.