Supporting Information for "Rapid Protein Global Fold Determination Using Ultrasparse Sampling, High-Dynamic Range Artifact Suppression, and Time-Shared NOESY" by Brian E. Coggins, Jonathan W. Werner-Allen, Anthony Yan, and Pei Zhou

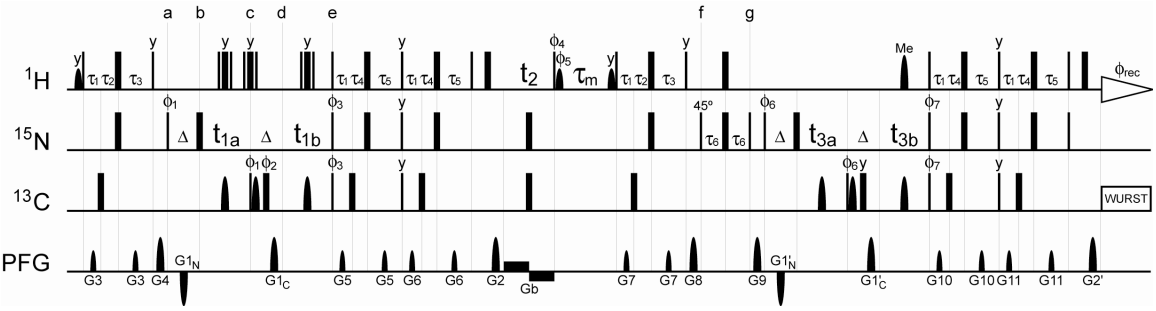

Figure S1. Pulse sequence for the 4-D TS NOESY experiment. Narrow and wide bars represent 90° and 180° pulses, respectively. All pulses are applied along the x-axis unless noted otherwise. 90° selective water pulses are indicated by short, shaped bars. All shaped carbon pulses are 281 μs off-resonance Isnob2 pulses[1] (centered at 175 ppm with a bandwidth of 30 ppm) that refocus carbonyl coupling during nitrogen chemical shift evolution. The delays are $\tau_1 = 1.7$ ms $\approx 1/4J_{CH}$, $\tau_2 = 0.7$ ms $\approx 1/4J_{NH} - 1/4J_{CH}$, $\tau_3 = 2.4$ ms $\approx 1/4J_{NH}$, $\tau_4 = 1.05$ ms $\approx 1/4J_{NH} - 1/4J_{CH}$, $\tau_5 = 2.75$ ms $\approx 1/4J_{NH}$, $\tau_6 = 1.34$ ms $\approx 1/8J_{NH}$, $\tau_m =$ 200 ms, and $\Delta = 200$ μs. Time increments are set to $\Delta t_{1a} = 1/sw_N - 1/sw_C$, $\Delta t_{1b} = 1/sw_C$, $\Delta t_2 = 1/sw_H$, $\Delta t_{3a} = 1/sw_N - 1/sw_C$, and $\Delta t_{3b} = 1/sw_C$. The proton inversion pulse at point c is required to refocus J_{NH} scalar coupling in the two Δ delays (points a-b and c-d), while proton decoupling during t_{1a} (points b-c) and t_{1b} (points d-e) evolution is achieved by the proton inversion pulses centered in these periods. The shaped bar marked 'Me' represents a 587 μs off-resonance G3 inversion pulse[2] (centered at -2.0 ppm with a bandwidth of 8.0 ppm) that selectively refocuses methyl proton coherence during t_{3b} without affecting water and amide signals. The ^{15}N carrier frequency is shifted 45 Hz downfield during the NOE mixing period to re-center nitrogen signals on the TROSY component. The spin-state-selective element between points f and g is used for active suppression of the ^1H-^{15}N anti-TROSY component.[3] Carbon decoupling during acquisition is achieved by using a WURST-40 sequence[4] with a field strength of 8.0 kHz. Phase cycling is $\phi_1 = [2x, 2(-x)]$, $\phi_2 = [x]$, $\phi_3 = [x]$, $\phi_4 = [-x]$, $\phi_5 = [x]$, $\phi_6 = [x, -x]$, $\phi_7 = [x]$, $\phi_{rec} = [x, -x, -x, x]$. Inversion of ϕ_4 (and ϕ_5 for water suppression) at even numbered lattice points in t_2 introduces a frequency shift of $sw_2/2$ to the H1 dimension in order to center the amide signals while leaving the transmitter frequency on water. Axial peaks are removed by setting ($\phi_1 + 180°$, $\phi_{rec} + 180°$) and ($\phi_6 + 180°$, $\phi_{rec} + 180$) at even numbered lattice points in F1 and F3, respectively. Hypercomplex data collection for the two time-shared, sensitivity-enhanced coherence transfers requires inversion of ϕ_3, ϕ_{rec}, and G2 for the F1 dimension and inversion of ϕ_7 and G2' for the F3 dimension. Cosine-sine selection for the F1/ F2 dimensions is controlled by incrementing phase ϕ_4 (and phase ϕ_5

for water suppression). For the gradient selection of nitrogen and carbon coherence pathways, nitrogen single quantum coherence is encoded with the sum of $G1_N$ and $G1_C$ (or $G1'_N$ and $G1'_C$ in the second transfer) whereas carbon single quantum coherence is encoded only by $G1_C$. Therefore, the duration of these gradients and the decoding gradient G2 are set such that $\tau_{G1C} = 4\tau_{G2}$ and $\tau_{G1C} + \tau_{G1N} = 10\tau_{G2}$, while the field strengths are optimized empirically, with the $G1_C$ gradient calibrated first. Gradient durations and field strengths are $G1_N$ = (1.2 ms, -18.34 G/cm), $G1_C$ = (0.8 ms, 18.51 G/cm), G2 = (0.2 ms, 18.38 G/cm), $G1'_N$ = (1.2 ms, -20.38 G/cm), $G1'_C$ = (0.8 ms, 20.56 G/cm), G2' = (0.2 ms, 20.42 G/cm), G3 = (0.5 ms, 11.64 G/cm), G4 = (1 ms, -18.99 G/cm), G5 = (0.5 ms, 17.77 G/cm), G6 = (0.5 ms, 15.72 G/cm), G7 = (0.5 ms, 10.01 G/cm), G8 = (0.7 ms, 19.81 G/cm), G9 = (1 ms, -18.99 G/cm), G10 = (0.5 ms, 10.41 G/cm), and G11 = (0.5 ms, 14.09 G/cm). A small refocusing gradient is applied during t_2 (Gb) to suppress water radiation damping. In order to separate NOESY pathways originating from methyl and amide protons during data processing, two sets of FIDs are collected for each set of ($t_{1a} + t_{1b}$, t_2, $t_{3a} + t_{3b}$) delays, with ϕ_2 = [y] in the second set of FIDs to selectively invert methyl signals during the first coherence transfer.[5]

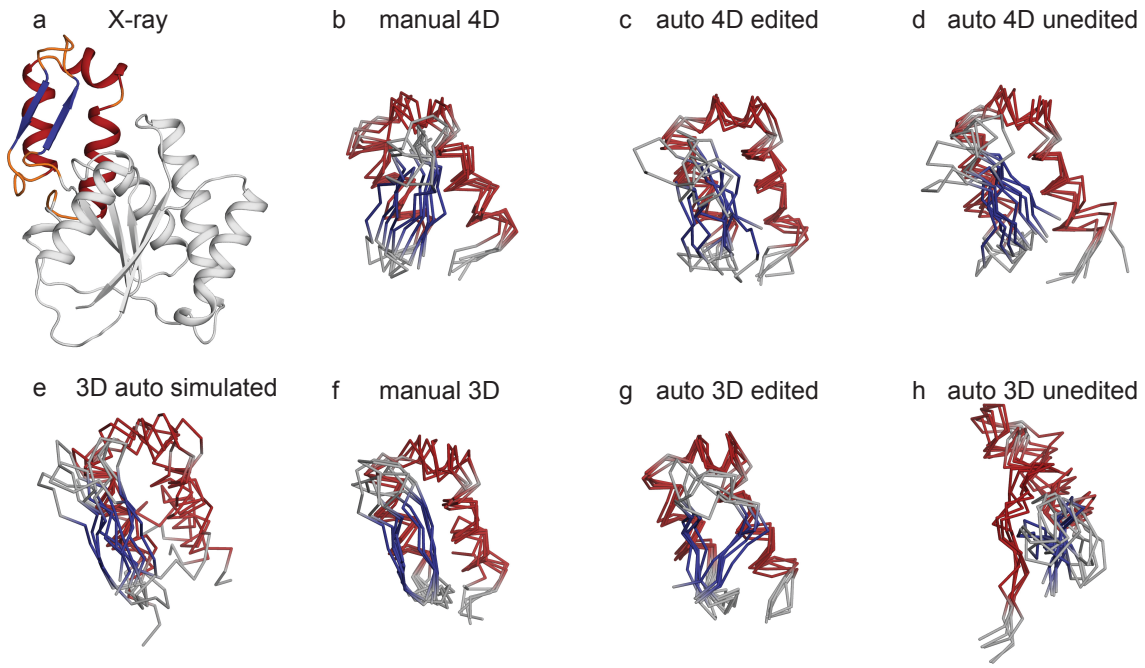

Figure S2. Alignment of the substrate-binding subdomain of Ssu72, which is not constrained to the main phosphatase domain by the TS NOESY data, in ensembles from CYANA structure calculations. The inputs were manually-assigned and manually-edited peaks (b, f), auto-assigned and manually-edited peaks (c, g), and auto-assigned and unedited peaks (d, h). In (b-d), peak lists are from SCRUB-processed 4-D TS spectra. In (f-h), peak lists are from conventional 3-D TS spectra. In (e), calculations are based on simulated 3-D peak lists derived from 4-D spectra. The reference crystal structure (PDB code 3FDF) is shown in (a).

		manual edited 4-D	auto edited 4-D	auto unedited 4-D	auto simulated 3-D
Peak Assignment Statistics [a]					
all spectra	*correct – all* [b]	1818	1778 (233)	n/a	1529 (674)
	correct – long-range [b]	698	678 (126)		485 (277)
	unassigned	0	32		60
	inconsistent [c]	0	8		229 (102)
amide-amide	*correct – all* [b]	879	853 (16)	n/a	738 (222)
	correct – long-range [b]	271	261 (8)		175 (64)
	unassigned	0	25		39
	inconsistent [c]	0	1		102 (51)
methyl-methyl	*correct – all* [b]	346	341 (142)	n/a	252 (170)
	correct – long-range [b]	228	224 (97)		156 (116)
	unassigned	0	0		2
	inconsistent [c]	0	5		92 (51)
amide-methyl	*correct – all* [b]	330	323 (30)	n/a	300 (135)
	correct – long-range [b]	114	110 (11)		90 (52)
	unassigned	0	5		9
	inconsistent [c]	0	2		21
methyl-amide	*correct – all* [b]	263	261 (45)	n/a	239 (147)
	correct – long-range [b]	85	83 (10)		64 (45)
	unassigned	0	2		10
	inconsistent [c]	0	0		14
Ensemble Convergence and Accuracy [d]					
Mean $RMSD_{bb,core}$ (Å)		0.904 ± 0.084	0.958 ± 0.081	1.193 ± 0.118	3.128 ± 1.272
Mean $RMSD_{bb,full}$ (Å)		1.340 ± 0.109	1.774 ± 0.272	2.500 ± 0.525	9.579 ± 2.648
Mean $bias_{bb,core}$ (Å)		1.634 ± 0.075	1.836 ± 0.300	2.665 ± 0.611	8.731 ± 3.430
Mean $bias_{bb,full}$ (Å)		2.455 ± 0.084	3.331 ± 0.185	5.039 ± 1.374	16.689 ± 2.177

[a] Assignment statistics are presented for a single CYANA structure calculation that produced an ensemble with the median $RMSD_{bb,core}$ out of five independent calculations. The unedited peak lists were not assigned manually, and therefore the accuracy of auto-assignment by CYANA cannot be assessed.

[b] Numbers in parentheses indicate peaks with ambiguous assignments.

[c] 'Inconsistent' denotes peaks with automated assignments that differ from the manual ones; however, for the 4-D peak lists, the automated assignment is also compatible with the reference crystal structure in all cases.

[d] Mean RMSD and bias were calculated with five ensembles – produced by independent CYANA runs and containing five structures each – over all non-hydrogen backbone atoms (bb) in the converged portion of either the full protein (full, residues 22-260) or residues in secondary structure elements (core). Bias represents the average pairwise RMSD to the reference crystal structure (PDB code 2ILI).

Table S1. Peak assignment and structure determination statistics for CYANA global fold calculations of HCA2 with peak lists from 4-D time-shared spectra.

		manual edited 3-D	auto edited 3-D	auto unedited 3-D
Peak Assignment Statistics [a]				
all spectra	*correct – all* [b]	2121	1659 (322)	n/a
	correct – long-range [b]	893	624 (149)	
	unassigned		348	
	inconsistent [c]		114	
amide-amide	*correct – all* [b]	898	689 (54)	n/a
	correct – long-range [b]	288	193 (18)	
	unassigned		196	
	inconsistent [c]		13	
methyl-methyl	*correct – all* [b]	271	230 (109)	n/a
	correct – long-range [b]	181	147 (78)	
	unassigned		21	
	inconsistent [c]		20	
amide-methyl	*correct – all* [b]	394	307 (69)	n/a
	correct – long-range [b]	154	100 (20)	
	unassigned		60	
	inconsistent [c]		27	
methyl-amide	*correct – all* [b]	558	433 (90)	n/a
	correct – long-range [b]	270	184 (33)	
	unassigned		71	
	inconsistent [c]		54	
Ensemble Convergence and Accuracy [d]				
Mean $RMSD_{bb,core}$ (Å)		0.895 ± 0.049	1.024 ± 0.116	1.429 ± 0.184
Mean $RMSD_{bb,full}$ (Å)		1.429 ± 0.120	2.310 ± 1.139	1.794 ± 0.113
Mean $bias_{bb,core}$ (Å)		1.752 ± 0.054	2.514 ± 0.239	13.222 ± 0.508
Mean $bias_{bb,full}$ (Å)		2.745 ± 0.096	6.040 ± 3.090	16.914 ± 1.169

[a] Assignment statistics are presented for a single CYANA structure calculation that produced an ensemble with the median $RMSD_{bb,core}$ out of five independent calculations. The unedited peak lists were not assigned manually, and therefore the accuracy of auto-assignment by CYANA cannot be assessed.

[b] Numbers in parentheses indicate peaks with ambiguous assignments.

[c] 'Inconsistent' denotes peaks with automated assignments that differ from the manual ones.

[d] Mean RMSD and bias were calculated with five ensembles – produced by independent CYANA runs and containing five structures each – over all non-hydrogen backbone atoms (bb) in the converged portion of either the full protein (full, residues 22-260) or residues in secondary structure elements (core). Bias represents the average pairwise RMSD to the reference crystal structure (PDB code 2ILI).

Table S2. Peak assignment and structure determination statistics for CYANA global fold calculations of HCA2 with peak lists from 3-D time-shared spectra.

		manual edited 4-D	auto edited 4-D	auto unedited 4-D	auto simulated 3-D
Peak Assignment Statistics [a]					
all spectra	*correct – all* [b]	1563	1517 (174)	n/a	1280 (621)
	correct – long-range [b]	490	461 (89)		296 (185)
	unassigned	0	28		27
	inconsistent [c]	0	18		256 (129)
amide-amide	*correct – all* [b]	703	692 (19)	n/a	618 (216)
	correct – long-range [b]	91	84 (0)		55 (21)
	unassigned	0	10		16
	inconsistent [c]	0	1		69 (50)
methyl-methyl	*correct – all* [b]	307	300 (75)	n/a	181 (128)
	correct – long-range [b]	201	195 (55)		110 (85)
	unassigned	0	1		2
	inconsistent [c]	0	6		124 (79)
amide-methyl	*correct – all* [b]	321	302 (21)	n/a	275 (131)
	correct – long-range [b]	123	114 (9)		81 (43)
	unassigned	0	16		9
	inconsistent [c]	0	3		37
methyl-amide	*correct – all* [b]	232	223 (59)	n/a	206 (146)
	correct – long-range [b]	75	68 (25)		50 (36)
	unassigned	0	1		0
	inconsistent [c]	0	8		26
Ensemble Convergence and Accuracy [d]					
Mean $RMSD_{bb,core}$ (Å)		0.923 ± 0.128	1.009 ± 0.158	0.996 ± 0.098	8.784 ± 1.305
Mean $RMSD_{bb,full}$ (Å)		1.291 ± 0.095	1.296 ± 0.114	1.191 ± 0.068	9.242 ± 1.514
Mean $bias_{bb,core}$ (Å)		2.076 ± 0.126	2.926 ± 0.262	3.576 ± 0.226	25.461 ± 4.462
Mean $bias_{bb,full}$ (Å)		2.413 ± 0.133	3.107 ± 0.198	4.176 ± 0.274	25.252 ± 4.148

[a] Assignment statistics are presented for a single CYANA structure calculation that produced an ensemble with the median $RMSD_{bb,core}$ out of five independent calculations. The unedited peak lists were not assigned manually, and therefore the accuracy of auto-assignment by CYANA cannot be assessed.

[b] Numbers in parentheses indicate peaks with ambiguous assignments.

[c] 'Inconsistent' denotes peaks with automated assignments that differ from the manual ones; however, for the 4-D peak lists, the automated assignment is also compatible with the reference crystal structure in all cases.

[d] Mean RMSD and bias were calculated with five ensembles – produced by independent CYANA runs and containing five structures each – over all non-hydrogen backbone atoms (bb) in the converged portion of either the full protein (full, residues 5-37, 97-195) or residues in secondary structure elements (core). Bias represents the average pairwise RMSD to the reference crystal structure (PDB code 3FDF).

Table S3. Peak assignment and structure determination statistics for CYANA global fold calculations of Ssu72 with peak lists from 4-D time-shared spectra.

		manual edited 3-D	auto edited 3-D	auto unedited 3-D
Peak Assignment Statistics [a]				
all spectra	*correct – all* [b]	1923	1537 (245)	n/a
	correct – long-range [b]	724	526 (127)	
	unassigned		267	
	inconsistent [c]		119	
amide-amide	*correct – all* [b]	825	682 (34)	n/a
	correct – long-range [b]	171	121 (7)	
	unassigned		127	
	inconsistent [c]		16	
methyl-methyl	*correct – all* [b]	193	158 (62)	n/a
	correct – long-range [b]	120	102 (46)	
	unassigned		17	
	inconsistent [c]		18	
amide-methyl	*correct – all* [b]	440	345 (78)	n/a
	correct – long-range [b]	202	145 (40)	
	unassigned		56	
	inconsistent [c]		39	
methyl-amide	*correct – all* [b]	465	352 (71)	n/a
	correct – long-range [b]	231	158 (34)	
	unassigned		67	
	inconsistent [c]		46	
Ensemble Convergence and Accuracy [d]				
Mean $RMSD_{bb,core}$ (Å)		0.668 ± 0.054	1.090 ± 0.117	1.307 ± 0.174
Mean $RMSD_{bb,full}$ (Å)		0.925 ± 0.079	1.372 ± 0.191	1.428 ± 0.176
Mean $bias_{bb,core}$ (Å)		1.856 ± 0.071	4.361 ± 1.847	12.300 ± 0.517
Mean $bias_{bb,full}$ (Å)		2.160 ± 0.079	5.214 ± 2.055	13.691 ± 0.553

[a] Assignment statistics are presented for a single CYANA structure calculation that produced an ensemble with the median $RMSD_{bb,core}$ out of five independent calculations. The unedited peak lists were not assigned manually, and therefore the accuracy of auto-assignment by CYANA cannot be assessed.

[b] Numbers in parentheses indicate peaks with ambiguous assignments.

[c] 'Inconsistent' denotes peaks with automated assignments that differ from the manual ones.

[d] Mean RMSD and bias were calculated with five ensembles – produced by independent CYANA runs and containing five structures each – over all non-hydrogen backbone atoms (bb) in the converged portion of either the full protein (full, residues 5-37, 97-195) or residues in secondary structure elements (core). Bias represents the average pairwise RMSD to the reference crystal structure (PDB code 3FDF).

Table S4. Peak assignment and structure determination statistics for CYANA global fold calculations of Ssu72 with peak lists from 3-D time-shared spectra.

References

(1) Kupce, E.; Boyd, J.; Campbell, I. D. *J Magn Reson Ser B* **1995**, *106*, 300.
(2) Emsley, L.; Bodenhausen, G. *Chem Phys Lett* **1990**, *165*, 469.
(3) Yang, D.; Kay, L. E. *J Biomol NMR* **1999**, *13*, 3.
(4) Kupce, E.; Freeman, R. *J Magn Reson Ser A* **1995**, *115*, 273.
(5) Frueh, D. P.; Vosburg, D. A.; Walsh, C. T.; Wagner, G. *J Biomol NMR* **2006**, *34*, 31.