Electronic Supplementary Material (ESI) for Molecular BioSystems.

Electronic Supporting Information for

A pre-structured helix in the intrinsically disordered 4EBP1

Do-Hyoung Kim, Chewook Lee, Ye-Jin Cho, Si-Hyung Lee, Eun-Ji Cha, Ji-Eun Lim, T. Michael Sabo, Christian Griesinger, Donghan Lee, and Kyou-Hoon Han

METHODS

Plasmids and cell strains

The clone, eukaryotic translation initiation factor 4E (eIF4E) binding protein 1 (4EBP1) (GenBank accession no. BC004459) cloned into the pOTB7, was obtained from the 21C Human Gene Bank, Genome Research Center, KRIBB, Korea. The truncated 4EBP1 gene, coding for amino acids 49-118 of the 4EBP1 protein (BP49), was excised from pOTB7 and cloned into a pGEX-2T plasmid (GE Healthcare) using *Bam*HI and *Eco*RI restriction enzyme sites. Bacterial strain DH5α (RBC Bioscience) was used to prepare all double stranded DNAs and BL21(DE3) cells (Novagen) were used for BP49 expression.

Protein purification

Cells transformed with pGEX-2T plasmid containing BP49 were grown overnight on LB agar plates supplemented with 50 μg/mL ampicillin. A single colony was transferred into LB media containing ampicillin and the cells were grown until it reached an OD_{600} of 0.6. Protein expression was induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The cells were then further cultivated at 20°C for 16 hrs. The ^{15}N-labeled or $^{13}C/^{15}N$-labeled BP49 was grown in M9 minimal media where the sole nitrogen source is ^{15}N labeled ammonium chloride and the sole carbon source is ^{13}C glucose, respectively. The cells were harvested by centrifugation and the bacterial pellet was resuspended in PBS buffer containing 0.1% Triton X-100 and 0.5 mg/mL lysozyme. After incubation on ice for 30 min, the cells were lysed by three subsequent freeze-thaw cycles. Genomic DNA was digested by 5 μg/mL recombinant DNAse I (TaKaRa) for 20 min at 37°C. The cells were sonicated using an Ultra Cell TM (Sonics and Materials) and the pellet was separated by centrifugation (25,000 g for 25 min at 4°C). The supernatant was collected and filtered through a 0.45 μm syringe filter (Corning, SFCA membrane). The supernatant containing the GST-tagged protein was loaded onto the Glutathione-sepharose column (GE Healthcare) and eluted with a buffer containing 50 mM Tris-HCl (pH 8.0) and 10 mM reduced glutathione. The GST-tagged BP49 protein containing a thrombin

recognition site was cleaved by thrombin for 6 hrs at 37°C. When the reaction was terminated, thrombin was effectively inhibited by 1 mM phenylmethane-sulfonylfluoride (PMSF). The fraction containing BP49 protein was applied to a Sephacryl S-200 column (2.5 × 120 cm; GE Healthcare) equilibrated with a elution buffer (25 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2 mM DTT, and 0.02% NaN_3). The molecular weights of the purified proteins were confirmed by MALDI-TOF mass spectrometry. Protein concentration was determined by measuring the absorbance using a UV spectrophotometer (GE Healthcare) in a 1 cm path length cell.

NMR spectroscopy

For NMR experiments stable-isotope labeled (^{15}N and $^{15}N/^{13}C$) protein samples with a concentration of ~0.5 mM were prepared in 90% H_2O/10% ^2H_2O containing 20 mM sodium acetate, pH 6.3, 50 mM NaCl, 1 mM DTT, 0.1 mM PMSF, and 0.01 mM EDTA. NMR spectra were acquired using a Varian Unity INOVA 600 MHz or a Bruker Avance II 900 MHz equipped with a cryogenic probe. Sequence-specific resonance assignment of BP49 for 65 out of 70 residues was obtained using standard multidimensional double- and triple-resonance NMR techniques as previously described[1]. To achieve sequence-specific backbone and side chain assignment of all aliphatic residues of BP49, 2D ^1H-^{15}N HSQC, 3D HNCACB, 3D CBCA(CO)NH, 3D CC(CO)NH, 3D HNCO, 3D HN(CA)CO, 3D HNCA, 3D HBHA(CO)NH, 3D ^{15}N-edited TOCSY-HSQC, and ^{15}N-edited NOESY-HSQC (τ_{mix} = 80-150ms) were obtained at 5°C. The sequence-corrected random coil chemical-shift values of Schwarzinger *et al*.,[2] were used to calculate the secondary structure-related chemical shift deviations of Hα and Cα. The secondary structure propensity program (SSP)[3] was used to calculate the propensity for secondary structures in BP49. Temperature coefficients for the backbone amide protons (Δ_{NH}) were calculated using the ^1H resonance assignments obtained at three temperatures (5°C, 10°C, and 15°C). The ^{15}N T_1 relaxation times were measured from spectra recorded with eight relaxation delays (20, 40, 80, 160, 320, 640, 1,280, and 2,560 ms) and the ^{15}N T_2 relaxation times were measured from spectra recorded using a CPMG sequence with eight relaxation delays (10, 30, 50, 70, 90, 130, 190, and 250 ms) according to the published procedure[4]. The ^1H-^{15}N heteronuclear steady-state NOEs

were measured from a pair of spectra recorded with and without a proton pre-saturation. All data were processed and analyzed on a Red Hat Linux version using a Varian VnmrJ, nmrPipe[5] and Sparky software. The spectral density mapping[6] was performed with the previously described ^{15}N T_1, T_2, and heteronuclear stead-state NOE data using a suite of Mathematica notebooks[7].

In order to measure RDCs, BP49 (0.1 mM) was added to a suspension of 10 mg/ml Pf-1 phage (ASLA Ltd., Riga, Lativa) in the NMR sample buffer[8]. The 2D-IPAP-^{15}N,^{1}H-HSQC experiment was utilized for both the aligned protein and the isotropic reference samples[9]. These experiments were recorded at a temperature of 278 K. The isotropic reference experiment was recorded on a Bruker Avance I 600 MHz spectrometer equipped with a cryogenic probe. Both the in-phase and anti-phase experiments were measured with 512 and 465 complex points in the direct (t_2) and indirect (t_1) dimensions, respectively, and 16 scans per t_1 increment. The $t_{1,max}$ and $t_{2,max}$ were 239 ms and 71.2 ms, respectively. The aligned protein sample was recorded on a Bruker Avance I 900 MHz spectrometer equipped with a cryogenic probe. Both the in-phase and anti-phase experiments were measured with 512 and 434 complex points in the direct (t_2) and indirect (t_1) dimensions, respectively, and 112 scans per t_1 increment. The $t_{1,max}$ and $t_{2,max}$ were 148 ms and 47.5 ms, respectively. Frequency discrimination in the indirectly detected dimension was achieved with the States-TPPI scheme[10].

All time domain data were processed in the same manner with NMRPipe software[5]. The data were zero-filled to 4 and 8 k in t_1 and t_2, respectively. A sine-bell window function was applied in both dimensions and a polynomial baseline correction was applied in the direct dimension. The NH J- and (RDC+J)-coupling constants were extracted using NMRPipe. One bond ^{15}N,^{1}H RDCs were derived from the difference in coupling between the aligned and the isotropic samples. The error in the peak position was determined by dividing the peak's line-width by the peak's signal-to-noise ratio[11]. After propagating the error in the peak position, RDC errors ranged from 0.07 to 3.4 Hz with a mean of 0.45 ± 0.60 Hz (see **Fig. 2h**).

Flexible Meccano (FM) calculations

The details of the FM algorithm have been previously discussed. Following the protocol outlined by Blackledge and co-workers[12], we have calculated ensembles of 10,000 BP49 conformers containing all possible combinations of helices from 4 to 13 residues in length spanning amino acids 51 to 63. Each ensemble is populated at 100% with the specified helix location and length. In addition to the helical ensembles, a 70,000 member unfolded ensemble of BP49 was also generated. From the FM generated ensembles of BP49, RDCs where back calculated and used to fit for the population of the given helix in the BP49 conformational ensemble. All possible combinations of BP49 ensembles comprising of n = 0, 1, 2, and 3 helices were compared to the measured RDCs utilizing the following equations for fitting:

$$D_{ij,eff} = \sum_{k=1,n} p_k D_{ij}^k + (1 - \sum_{k=1,n} p_k) D_{ij}^U \quad (1)$$

where p_k are the populations of the *n* helical conformers, for which D_{ij}^k are the individual predicted couplings between nuclei *i* and *j*, and D_{ij}^U are the couplings from the unfolded state, which were calculated from the described ensemble in the unfolded state (n=0). Next, the measured RDCs where used to minimize the following target function:

$$\chi^2 = \sum (A * D_{ij,eff} - D_{ij,measured})^2 / \sigma_{ij}^2 \quad (2)$$

where σ_{ij} is the error in the measured RDC. The scaling factor *A* is needed to account for the uncertainty in the absolute strength of the alignment media. In principal, an ensemble of inter-nuclear vectors can be put into a unique principal axis system[13] where relative RDCs can be back calculated without any information regarding the alignment media. Since the absolute strength of the alignment media will affect all inter-nuclear vectors equally, the global parameter *A* is introduced in the fitting to account for the unknown magnitude of the alignment. Using the error in the measured RDCs, 1000 Monte Carlo simulations were run to determine the uncertainties in the populations. To test the significance in the fit of the data to more complex models, a standard F-test was implemented and the *p* value was determined. Finally, the Q-Factor was calculated with the formula:

$$Q = \sum \sqrt{(A * D_{ij,eff} - D_{ij,measured})^2 / D_{ij,measured}^2} \quad (3)$$

Supplementary Table 1 summarizes the results of the best fitting combinations of helices within residues 51-63 of BP49 and **Supplementary Figure 2** graphically illustrates the comparison of the back calculated RDCs to the experimentally measured RDC_S. Intensities of NOE peaks were calculated similar to equation (1)

$$NOE_{ij} \propto \sum_{k=1,n} p_k \left\langle \frac{1}{\left(d_{ij}^k\right)^6} \right\rangle + (1 - \sum_{k=1,n} p_k) \left\langle \frac{1}{\left(d_{ij}^U\right)^6} \right\rangle \quad (4)$$

where p_k are the populations of the *n* helical conformers determined by FM approach, for which d_{ij}^k and d_{ij}^U are the distances between nuclei *i* and *j*, within α-helix and the ensemble of the unfolded state (n=0), respectively. **Supplementary Table 2** compiles the results of the ratio between the intensities of a typical NOE from a helix (between (i) HN(i) and HN(i+2) or (ii) Hα(i) and HN(i+3),) and the observed NOE between Hα(i) and HN(i+1).

Replica exchange molecular dynamics (REMD) calculations

The number of amino acid residues in the BP49 protein used for wet experiments is 70. The same number of residues (49-118) was used for MD simulation. The initial structure of BP49 was built by a *tleap* module in AMBER10 MD program package with two termini capped by acetyl beginning and amine ending group, respectively[14]. The initial conformation of BP49 was set to a fully extended β-sheet with two backbone dihedral angles (phi, psi) of 180.0°. An all atom ff99SB force field[15] was used to assign the parameters, which resulted in a total protein size with 1083 atoms. A GB/SA model[16] was used to mimic the solvent effect (igb=5). The extended polypeptide was subjected to a short minimization of 1,000 steps comprising of initial 500 steps of steepest descent and subsequent 500 steps of conjugated gradient method. The resulting minimized structure was further heated till 300K and then equilibrated for 200 ps at the same temperature. The equilibrated structure was subjected to replica exchange molecular dynamics (REMD) simulation[17] with random seeds of

initial velocity using the AMBER10 program package. The SHAKE algorithm[18] was used to fix the length of the covalent bond involving hydrogen atoms, which allows an integration time step of 2 fs. The temperature range of sixteen replica in the REMD simulation was 300 ~ 515 K. Each replica was simulated for 10 ns (a total of 160 ns). Exchanges between two replica were performed every 1 ps during the simulation. Clustering analysis using an MMTSB toolset[19] based on the RMSD value (cut-off: 8.0 Å) was applied to each temperature trajectory. Choosing the realistic ensemble structures for the free-state BP49 and the ^{1}Hα chemical shifts for the representative structure in each cluster was calculated by Shift-X program[20]. The chemical shift values obtained from the calculation were compared with the experimental ^{1}Hα values using a Pearson's correlation (PC). Highly correlated structures (> 0.60 of PC coefficient) were obtained.

Supplementary Table S1. Best fitting of BP49 ensembles to the measured RDCs.

number of helices	helix range	helix population[a]	Q-Factor	$\chi^{13,b}$	p^c
1	57-62	13.2 ± 0.2 %	0.24	1958	
2	57-62	6.6 ± 0.2 %	0.18	312	0.04
	51-60	8.3 ± 0.1 %			
3	57-62	8.8 ± 0.2 %	0.12	112	0.75
	51-60	6.7 ± 0.1 %			
	51-54	9.3 ± 0.5 %			

[a]Population of the corresponding helix, with the remaining population being unfolded.

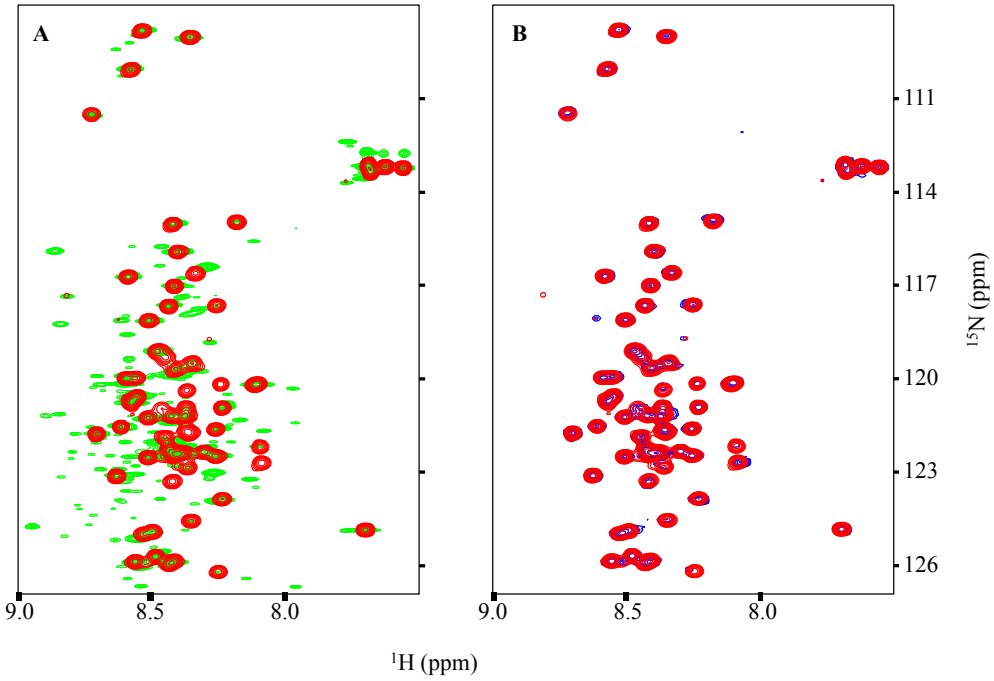
[b]Target function measured over 11 RDCs measured for residues 49-63.

[c]The significance of the improvement of the data fitting to a more complex model, determined using a standard F-test.

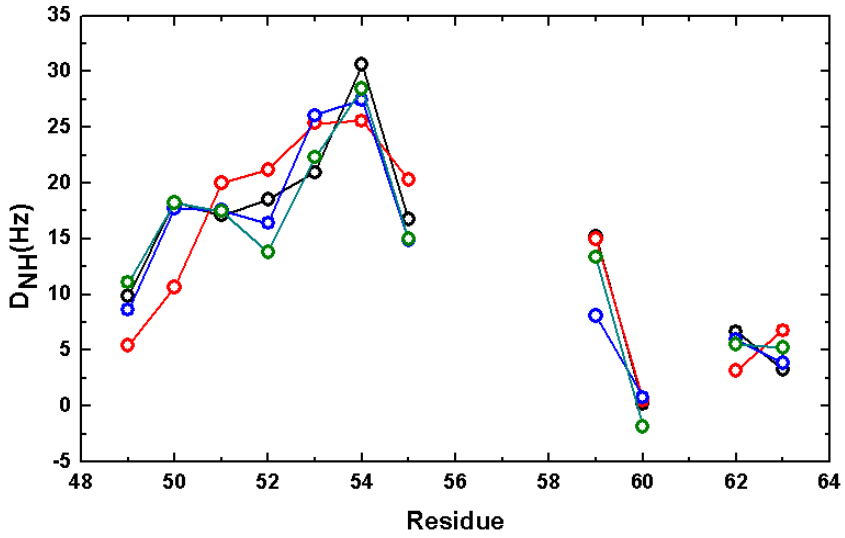
Supplementary Table S2. Expected NOE intensity for a pre-populated helix predicted from BP49 ensembles[a].

HN(i)-HN(i+2)	One helix ensemble (57-62) (%)	Two helix ensemble (57-62, 51-60) (%)	Hα(i)-HN(i+3)	One helix ensemble (57-62) (%)	Two helix ensemble (57-62, 51-60) (%)
51-53	0.9	1.4	51-54	0.4	2.0
52-54	0.5	0.8	52-55	0.3	1.2
53-55	0.8	1.1	53-56	0.3	1.3
54-56	1.5	1.8	54-57	0.8	1.9
55-57	1.9	2.3	55-58	1.6	2.7
56-58	2.1	2.3	56-59	4.4	4.2
57-59	1.7	1.8	**57-60**[b]	2.6	2.9
58-60	1.7	1.8	**58-61**	2.6	2.9
59-61	1.8	1.9	59-62	2.7	2.3
60-62	2.1	2.0			

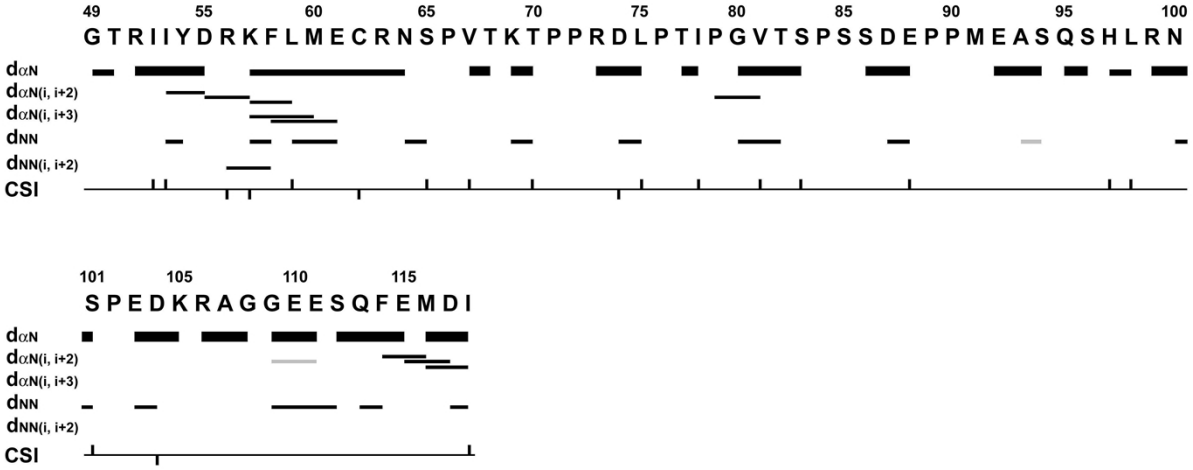
[a]Calculation performed according to equation (4) in Methods Section.
[b]Actual observed NOEs are in bold, all other expected NOEs are unobserved due to signal overlap.



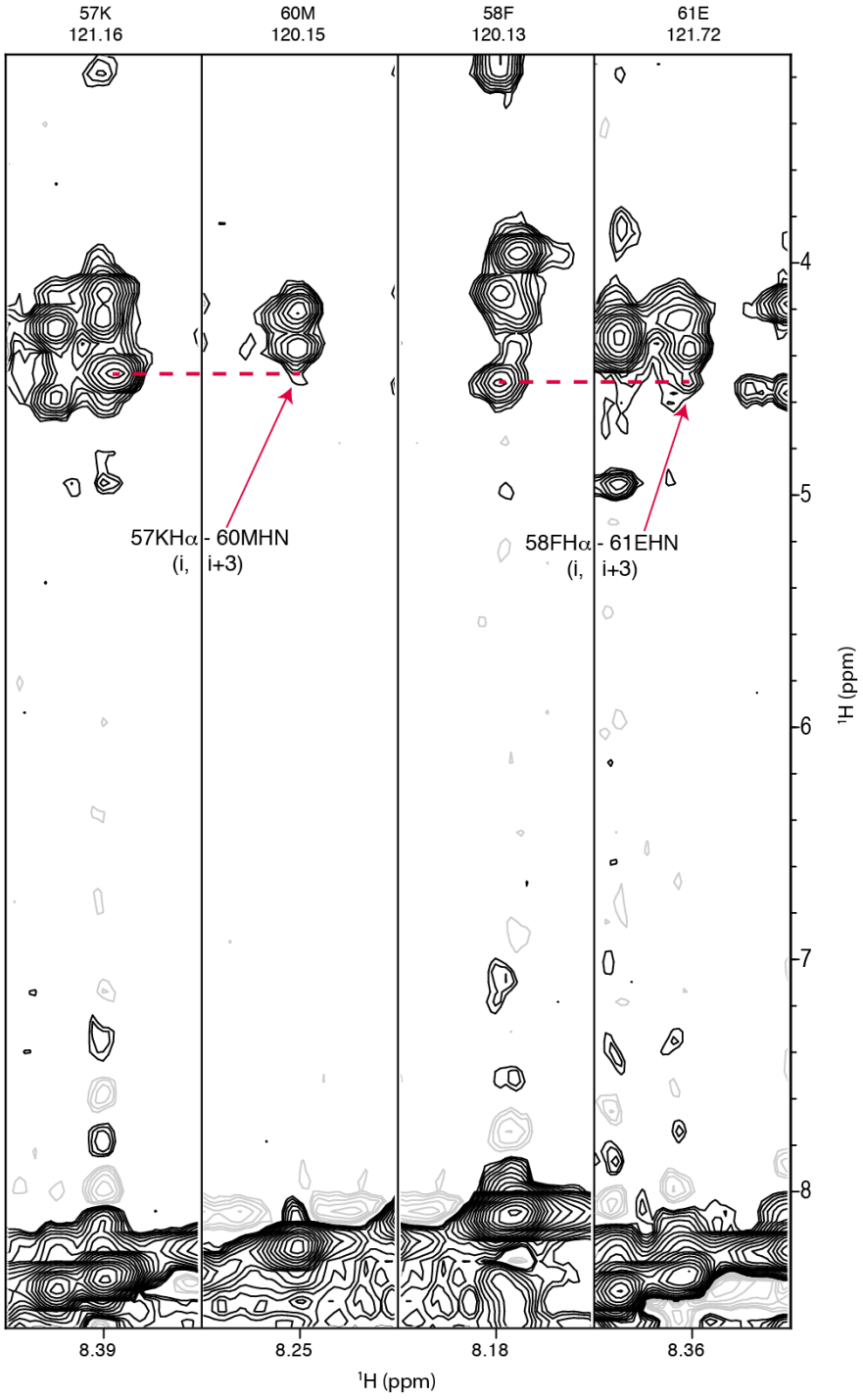
Supplementary Figure S1. ^{1}H-^{15}N heteronuclear HSQC spectra of BP49 under isotropic and anisotropic conditions. (A) Overlay of BP49 under isotropic conditions (red) and under anisotropic conditions (green) prepared with 10% dodecylpenta(ethylene glycol) and hexanol mixture in the NMR sample buffer[21]. (B) Overlay of BP49 under isotropic conditions (red) and under anisotropic conditions (blue) prepared with 10 mg/ml Pf-1 phage in the NMR sample buffer[8].



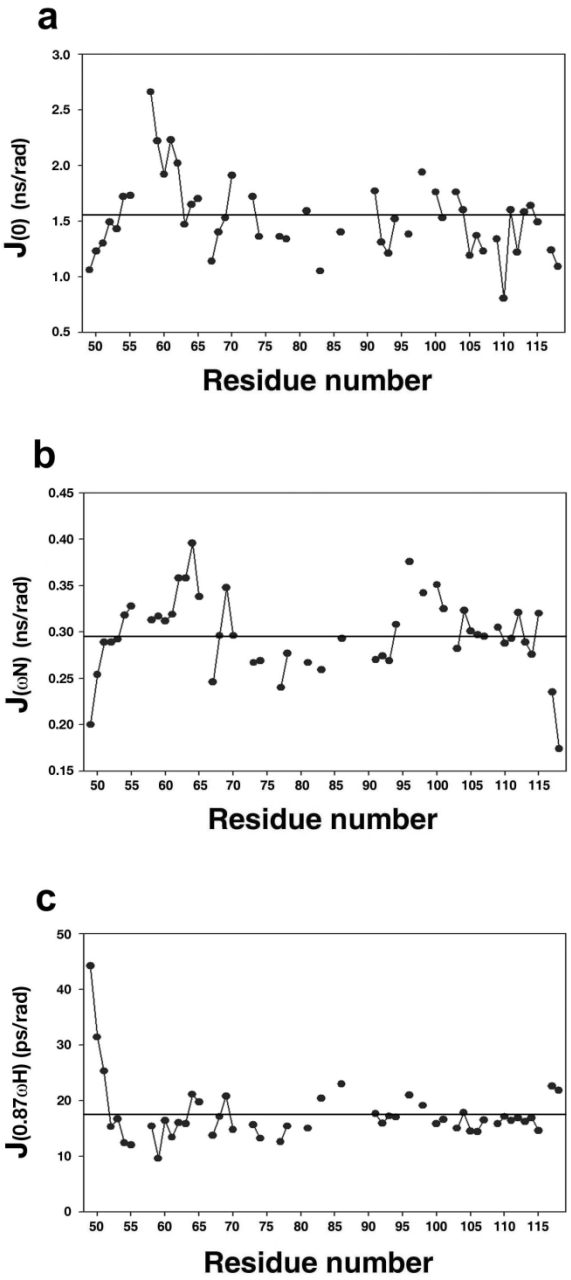
Supplementary Figure S2. The best fit of the experimental RDC data to the FM ensembles. Plot of the best fitting ensembles of n = 1 (red), 2 (blue), and 3 (green) helices to the measured RDCs (black). The best fitting ensemble containing one helix spanning residues 57-62 has a population of 13.2 ± 0.2 %, a Q-Factor of 0.24 and a χ^2 of 1958. The best fitting ensemble containing two helices spanning residues 57-62 and 51-60 has populations of 6.6 ± 0.2 % and 8.3 ± 0.1 %, respectively, a Q-Factor of 0.18, and a χ^2 of 312. The best fitting ensemble containing three helices spanning residues 57-62, 51-60, and 51-54 has populations of 8.8 ± 0.2 %, 6.7 ± 0.1 %, and 9.3 ± 0.5 % respectively, a Q-Factor of 0.12, and a χ^2 of 112.



Supplementary Figure S3. A summary of ^{1}H-^{1}H internuclear NOEs for BP49.



Supplementary Figure S4. Selected regions of the 3D ^{15}N-filtered NOESY measurement of BP49 displaying two possible NOEs, namely 57K(Hα) (i)-60M(HN) (i+3) and 58F(Hα) (i)-61E(HN) (i+3), which indicate the presence of a helix between residues 57-62.



Supplementary Figure S5. The reduced spectral density functions for BP49. (a) J(0), (b) J(ωN), (c) J(0.87ωH). The horizontal lines indicate an average. These indicate that the motion of PreSMo in BP49 is restricted compared to that of the rest of the molecules existing in a CU state. The transient nature of the helix PreSMo manifests also in the individual spectral densities that deviate from those of a stable helix

References for ESI

1. A. Bax and S. Grzesiek, *Acc. Chem. Res*., 1993, **26**, 131-138.

2. S. Schwarzinger, G. J. A. Kroon, T. R. Foss, J. Chung, P. E. Wright, and H. J. Dyson, *J. Am. Chem. Soc*., 2011, **123**, 2970-2978.

3. J. A. Marsh, V. K. Singh, Z. Jia, and J. D. Forman-Kay, *Protein Sci*., 2006, **15**, 2795-2804.

4. L. E. Kay, L. K. Nicholson, F. Delaglio, A. Bax, and D. A. Torchia, *J. Magn. Reson*., 1992, **97**, 359-375.

5. F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, *J. Biomol. NMR,* 1995, **6**, 277-293.

6. J. W. Peng and G. Wagner, *Biochemistry,* 1992, **31**, 8571-8586.

7. L. Spyracopoulos, *J. Biomol. NMR,* 2006, **36**, 215-224.

8. M. Zweckstetter and A. Bax, *J. Biomol. NMR,* 2001, **20**, 365-377.

9. M. Ottiger and A. Bax, *J. Am. Chem. Soc.,* 1998, **120**, 12334-12341.

10. D. Marion, M. Ikura, R. Tschudin, and A. Bax, *J. Magn. Reson.,* 1989, **85**, 393-399.

11. G. Kontaxis, G. M. Clore, and A. Bax, *J. Magn. Reson.,* 2000, **143**, 184-196.

12. M. R. Jensen, K. Houben, E. Lescop, L. Blanchard, R. W. J. Ruigrok, and M. Blackledge, *J. Am. Chem Soc.,* 2008, **130**, 8055-8061.

13. E. Meirovitch, D. Lee, K. F. Walter, and C. Griesinger, *J. Phys. Chem. B,* 2012, **116**, 6106-6117.

14. D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, and P. A. Kollman, AMBER 10 (2008), University of California, San Francisco.

15. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. L. Simmerling, *Proteins,* 2006, **65**, 712-725.

16. V. Tsui and D. A. Case, *Biopolymers,* 2000, **56**, 275-291.

17. A. Mitsutake, Y. Sugita, and Y. Okamoto, *Biopolymers (Peptide Sci.),* 2001, **60**, 96-123.

18. J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, *J. Comput. Phys.,* 1977, **23**, 327-341.

19. M. Feig, J. Karanicolas, and C. L. Brooks, III, *J. Mol. Graph Model.,* 2004, **22**, 377-395.

20. S. Neal, A. M. Nip, H. Zhang, and D. S. Wishart, *J. Biomol. NMR,* 2003, **26**, 215-240.

21. M. Ruckert and G. Otting, *J. Am. Chem. Soc.* 2000, **122**, 7793-7797.