Supplementary Material:

Strategy for complete NMR assignment of disordered proteins with highly repetitive sequences based on resolution-enhanced 5D experiments.

Veronika Motáčková[1], Jiří Nováček[1], Anna Zawadzka-Kazimierczuk[2], Krzysztof Kazimierczuk[2], Lukáš Žídek[1], Hana Šanderová[3], Libor Krásný[3], Wiktor Koźmiński[2], Vladimír Sklenář[1]

[1]Faculty of Science, National Centre for Biomolecular Research, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic.
[2]Faculty of Chemistry, University of Warsaw, Pasteura 1, 02 093 Warsaw, Poland.
[3]Laboratory of Molecular Genetics of Bacteria and Department of Bacteriology, Institute of Microbiology, Academy of Sciences of the Czech Republic, Vídeňská 1083, 142 20 Prague, Czech Republic.

Experimental setup

Conventional experiments

Conventional experiments were performed using the standard setup, described in the literature. Spectral parameters are listed in Table S1.

5D experiments with non-uniform sampling

HN(CA)CONH The high power 90° pulses with the length of 6.6 μs for ^{1}H, 13.5 μs for ^{13}C and 31.0 μs for ^{15}N were applied at 4.7 ppm for protons, at 120 ppm for ^{15}N, at 176.5 ppm for ^{13}C', and at 58.5 ppm for $^{13}C^{\alpha}$. The phase modulated *sinc* shape pulses were used for selective excitation or inversion of ^{13}C', $^{13}C^{\alpha}$, respectively. Decoupling of ^{1}H was achieved using WALTZ16 pulse train, whereas the adiabatic decoupling using ^{15}N WURST inversion pulses was used during the acquisition. The transfer of magnetisation was achieved using the INEPT transfer steps of the following lengths: $\Delta_{\mathrm{NH}} = 5.4$ ms, $\Delta_{\mathrm{NC}^{\alpha}} = 28.6$ ms, $\Delta_{\mathrm{C'C}^{\alpha}} = 9.1$ ms, $\Delta_{\mathrm{NC'}} = 28.0$ ms.
HabCabCONH Radio-frequency pulses were applied at 45.5 ppm for $^{13}C^{\alpha/\beta}$ nuclei, at 120.0 ppm for ^{15}N, and at 176.5 ppm for ^{13}C'. The carrier frequency for ^{1}H nuclei was first set in the aliphatic region at 3.15 ppm, and than shifted to 4.7 ppm. The high power 90° pulses for ^{1}H, ^{13}C, and ^{15}N were of the length of 6.5 μs, 13.5 μs, and 31.0 μs, respectively. The phase modulated *sinc* shape pulses were use for selective excitation of inversion of $^{13}C^{\alpha/\beta}$ and ^{13}C' nuclei. The WALTZ16 pulse train was used for proton decoupling during the magnetisation transfer, and adiabatic decoupling using ^{15}N WURST inversion pulses was used during acquisition. The evolution of the chemical shift in the ^{15}N dimension was applied in a semi-constant time mode. The following delays were used for the transfer of magnetisation via INEPT transfer steps: $\Delta_{\mathrm{H}^{\alpha/\beta}\mathrm{C}^{\alpha/\beta}} = 3.6$ ms, $\Delta_{\mathrm{C}^{\alpha}\mathrm{C}^{\beta}} = 14.3$ ms, $\Delta_{\mathrm{C'C}^{\alpha}} = 6.8$ ms, $\Delta_{\mathrm{NH}} = 5.4$ ms, and $\Delta_{\mathrm{NC'}} = 28.0$ ms.
HC(CC-TOCSY)CONH The high power 90° radio-frequency pulses with the length of 6.5 μs for ^{1}H, 13.5 μs for ^{13}C and 31.0 μs for ^{15}N were applied, at 120 ppm for ^{15}N, at 176.5 ppm for ^{13}C', and at 37.5 ppm for aliphatic carbon nuclei ($^{13}C^{\mathrm{aliph}}$). The carrier frequency for protons was originally set to 3.15 ppm and shifted to 4.7 ppm after the evolution of the chemical shifts of aliphatic protons. The phase modulated *sinc* shape pulses were use for selective excitation and inversion of $^{13}C^{\mathrm{aliph}}$, and ^{13}C' nuclei. The 14.5 ms DIPSI-3 spinlock with $\gamma B_2/2\pi$=7500 Hz was used for isotropic mixing of ^{13}C magnetisation. WALTZ16 pulse train was used for decoupling of protons during magnetisation transfer and adiabatic decoupling using ^{15}N WURST inversion pulses was used during acquisition. The evolution of the chemical shift in the ^{15}N dimension was applied in a semi-constant time mode. Where the magnetisation was transfered via INEPT transfer step, the following lengths of delays were used: $\Delta_{\mathrm{H}^{\mathrm{aliph}}\mathrm{C}^{\mathrm{aliph}}} = 3.6$ ms, $\Delta_{\mathrm{C'C}^{\alpha}} = 6.8$ ms, $\Delta_{\mathrm{NH}} = 5.4$ ms, and $\Delta_{\mathrm{NC'}} = 28.0$ ms.

Chemical shift statistics

Comparison of the chemical shift dispersion of RNA polymerase δ subunit with other disordered proteins is presented in Figures S1–S12. Fig. S1 shows C^{α} and C^{β} chemical shifts of aspartates, glutamates, lysines, and leucines of the δ subunit. Fig. S2 presents C^{α} chemical shifts of six proteins with the highest number of C^{α} chemical shifts within a range of 0.2 ppm out of 27 proteins described as disordered, unfolded, or unstructured in their BMRB entry titles. Figs. S3 to S12 show ^{1}H and ^{15}N chemical shifts of two proteins which spectra are presented in the paper and of additional eight disordered proteins randomly selected from BMRB database (first eight proteins of those described as disordered in their BMRB entry titles).

Table S1

Experiment	**Nucleus**			**Carrier ν (ppm)**			**Real Points**			**Scans**	**Spectral Width (Hz)**		
	X	Y	Z	X	Y	Z	X	Y	Z		X	Y	Z
H-N HSQC	^{1}H	^{15}N		4.744	116.310		2048	256		8	8389.262	1369.863	
HNCA	^{1}H	^{15}N	^{13}C	4.744	116.310	58.483	2048	48	128	8	8389.262	1369.863	3125.000
HN(CO)CA	^{1}H	^{15}N	^{13}C	4.744	116.310	58.483	2048	48	128	8	8389.262	1369.863	3125.000
HNCACB	^{1}H	^{15}N	^{13}C	4.744	116.310	44.004	1024	48	128	32	8389.262	1369.863	11111.111
CBCA(CO)NH	^{1}H	^{15}N	^{13}C	4.744	116.310	44.004	1024	48	128	16	8389.262	1369.863	11111.111
HNCO	^{1}H	^{15}N	^{13}C	4.744	116.310	176.331	2048	48	128	8	8389.262	1369.863	1666.667
HCCH-TOCSY	^{1}H	^{1}H	^{13}C	4.744	2.999	4.744	1024	128	64	16	8389.262	4194.618	3623.188
H-C NOESY	^{1}H	^{15}C	^{1}H	4.744	125.015	4.744	1536	48	240	16	8389.262	4545.455	7194.245
CON	^{13}C	^{15}N		175.006	116.310		1024	256		32	6009.615	1369.863	
CBCACON	^{13}C	^{13}C	^{15}N	175.006	44.044	116.310	1024	64	64	32	6009.615	11111.0	1369.823
CBCANCO	^{13}C	^{13}C	^{15}N	175.006	44.044	116.310	1024	64	64	32	6009.615	11111.0	1369.823

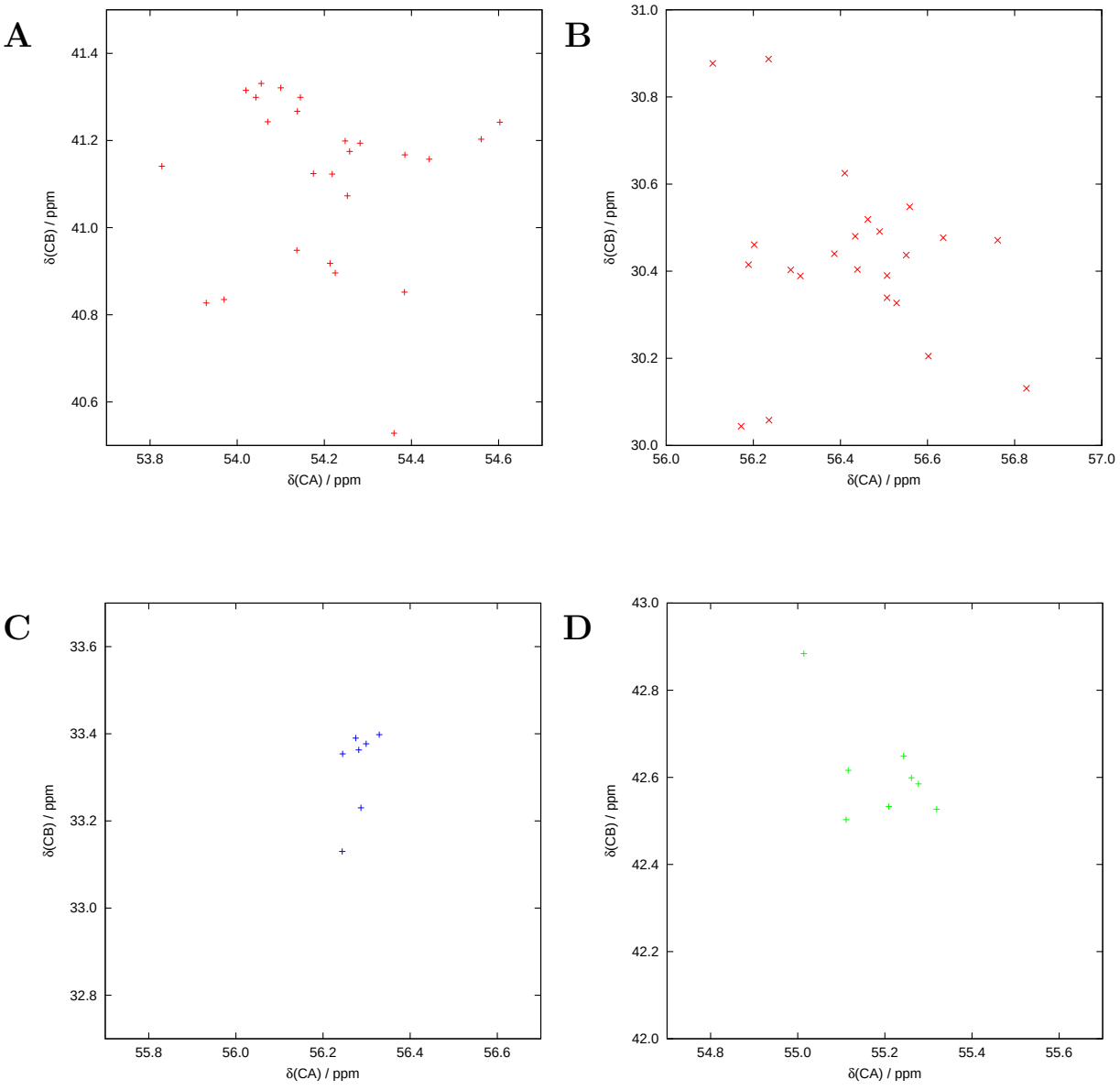

Figure S1: Correlation of $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts of aspartate (A), glutamate (B), lysine (C), and leucine (D) residues of the C-terminal domain of the RNA polymerase δ subunit. All $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts of these residues fall in the plotted 1 ppm wide regions, with the exception of the C-terminal Lys 173 ($^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts of 57.493 ppm and 33.900 ppm, respectively).

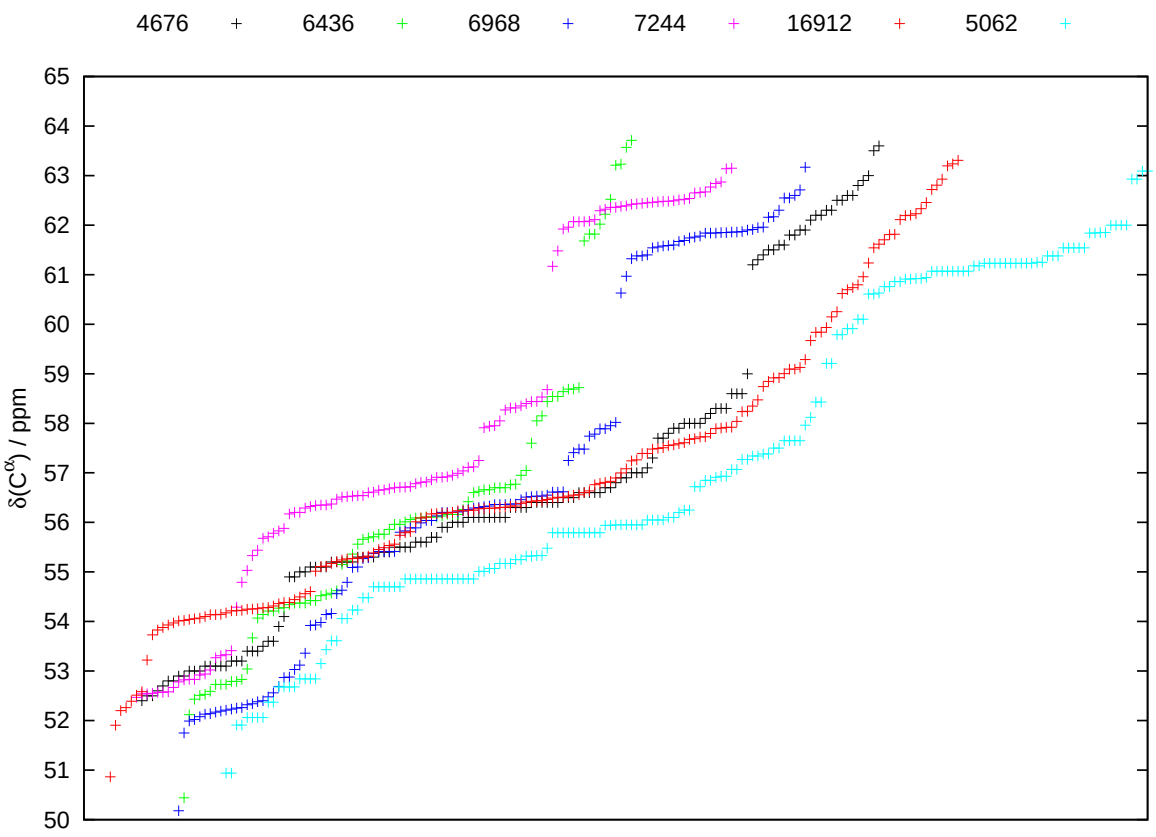

Figure S2: Chemical shifts of $^{13}C^{\alpha}$ nuclei of six disordered proteins with the highest number of $^{13}C^{\alpha}$ chemical shifts within the range of ± 0.2 ppm, sorted by their values. The BMRB Accession Numbers are listed above the plot, with the corresponding color coding. The RNA polymerase δ subunit, shown in red, exhibits particularly high degeneracy of the $^{13}C^{\alpha}$ chemical shifts.

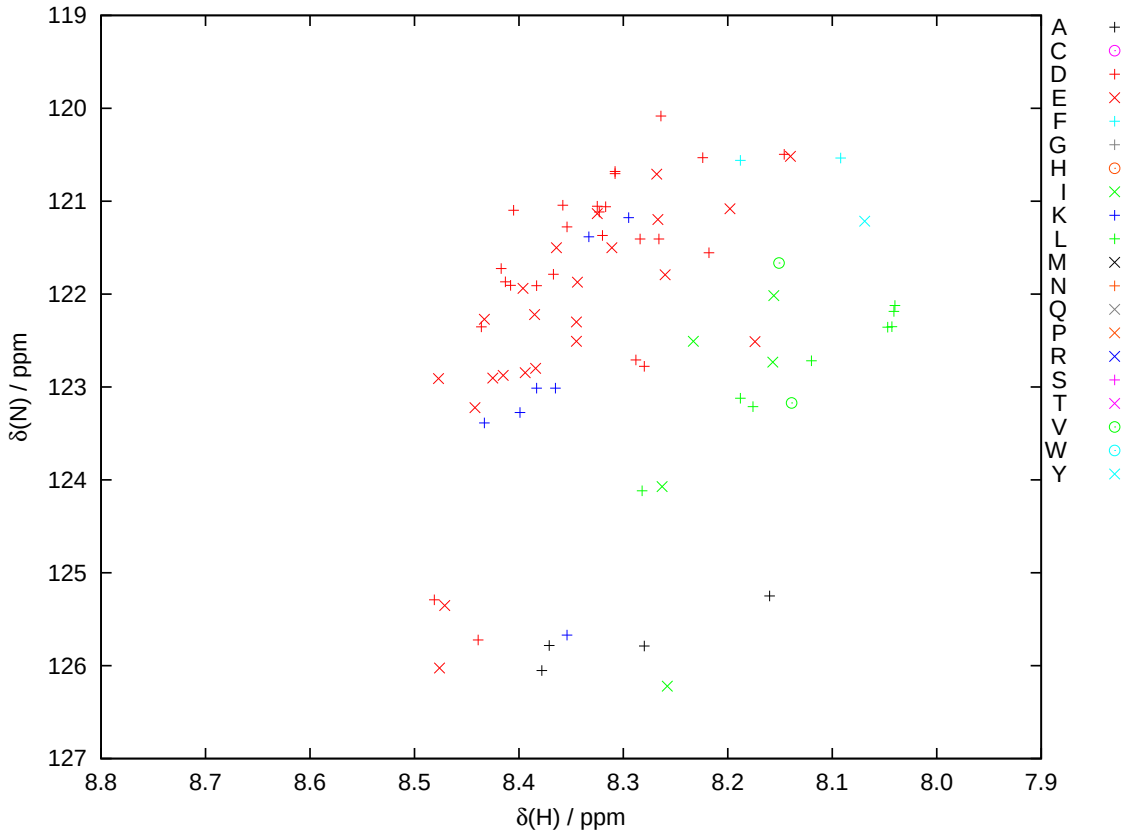

Figure S3: Correlation of ^{1}H and ^{15}N chemical shifts of the C-terminal domain of RNA polymerase δ subunit from *B. subtilis* (BMRB Accession Number 16912), studied in this work. The amino acid types are distinguished by point size and color, as indicated at the right-hand side.

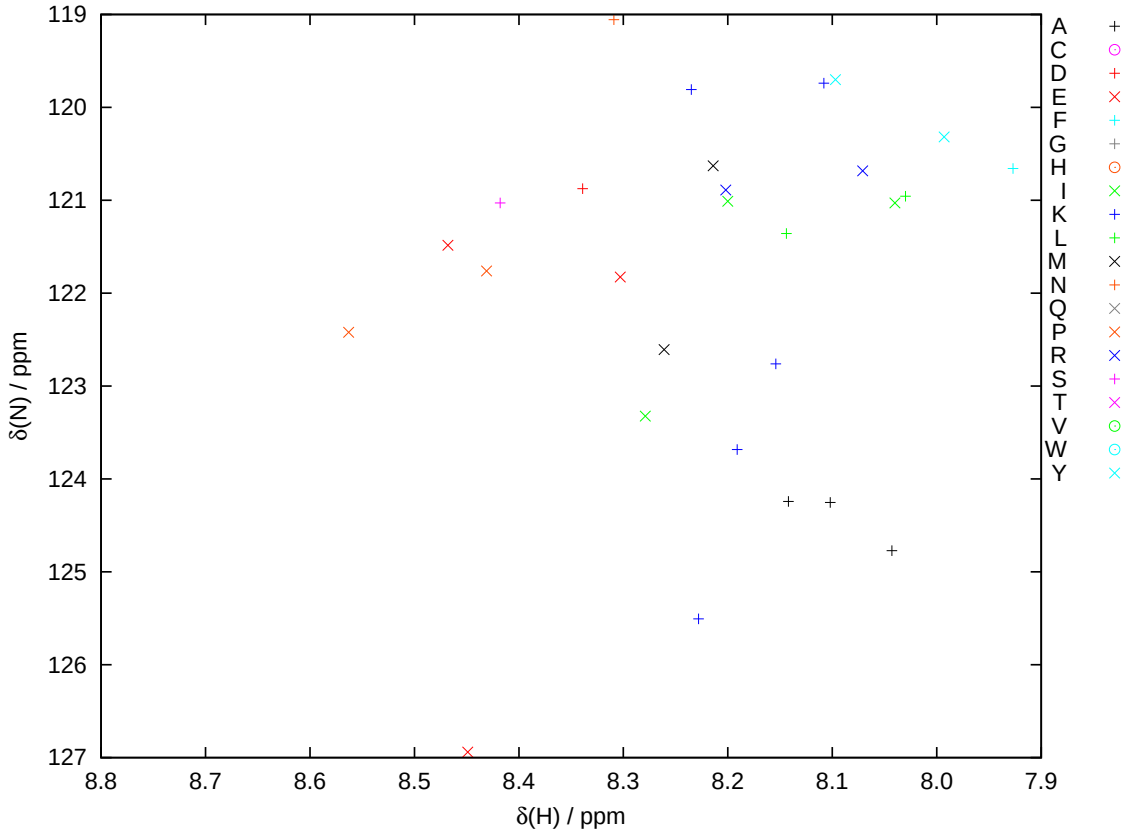

Figure S4: Correlation of ^{1}H and ^{15}N chemical shifts of the C-terminal domain of retroviral protease of murine intracisternal A-type particles (Motáčková et al. (2009) *Biomol. NMR Assign.* **3**, 261–264., BMRB Accession Number 16341). The amino acid types are distinguished by point size and color, as indicated at the right-hand side.

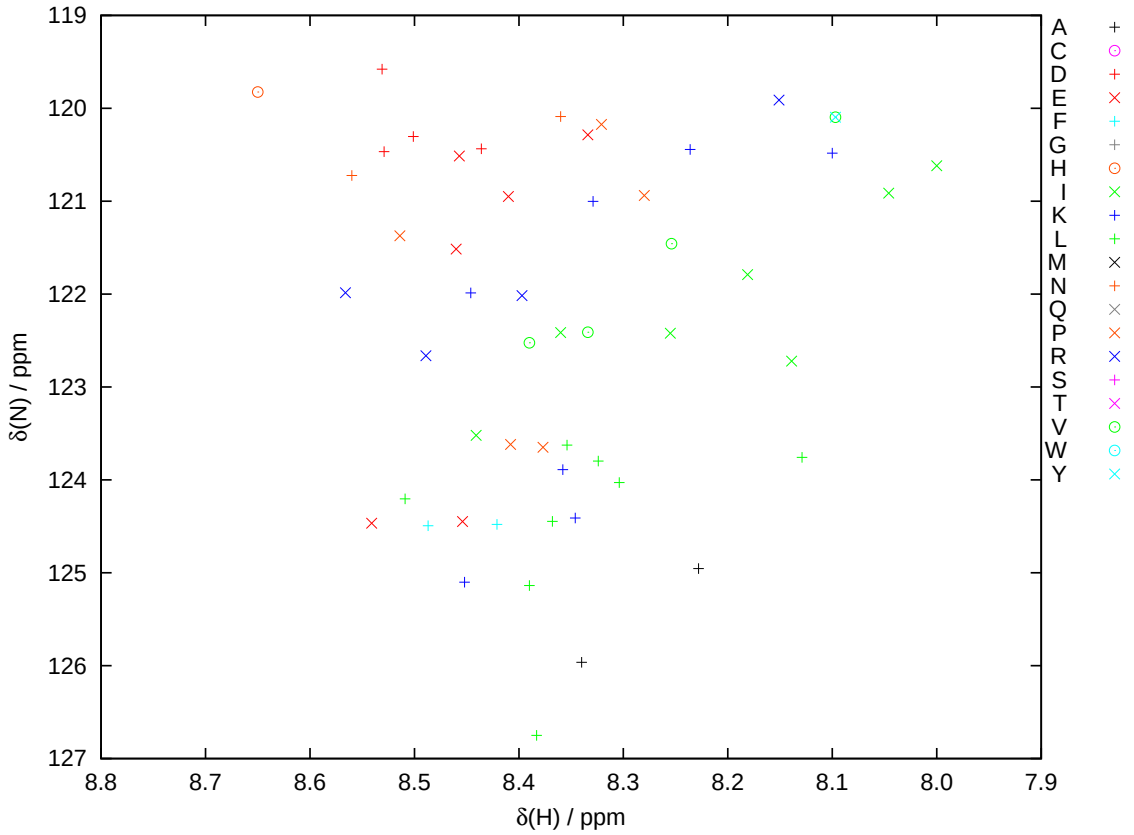

Figure S5: Correlation of ^{1}H and ^{15}N chemical shifts of denatured ubiquitin (Peti et al. (2001) *J. Biomol. NMR* **19**, 153–165., BMRB Accession Number 4375). The amino acid types are distinguished by point size and color, as indicated at the right-hand side.

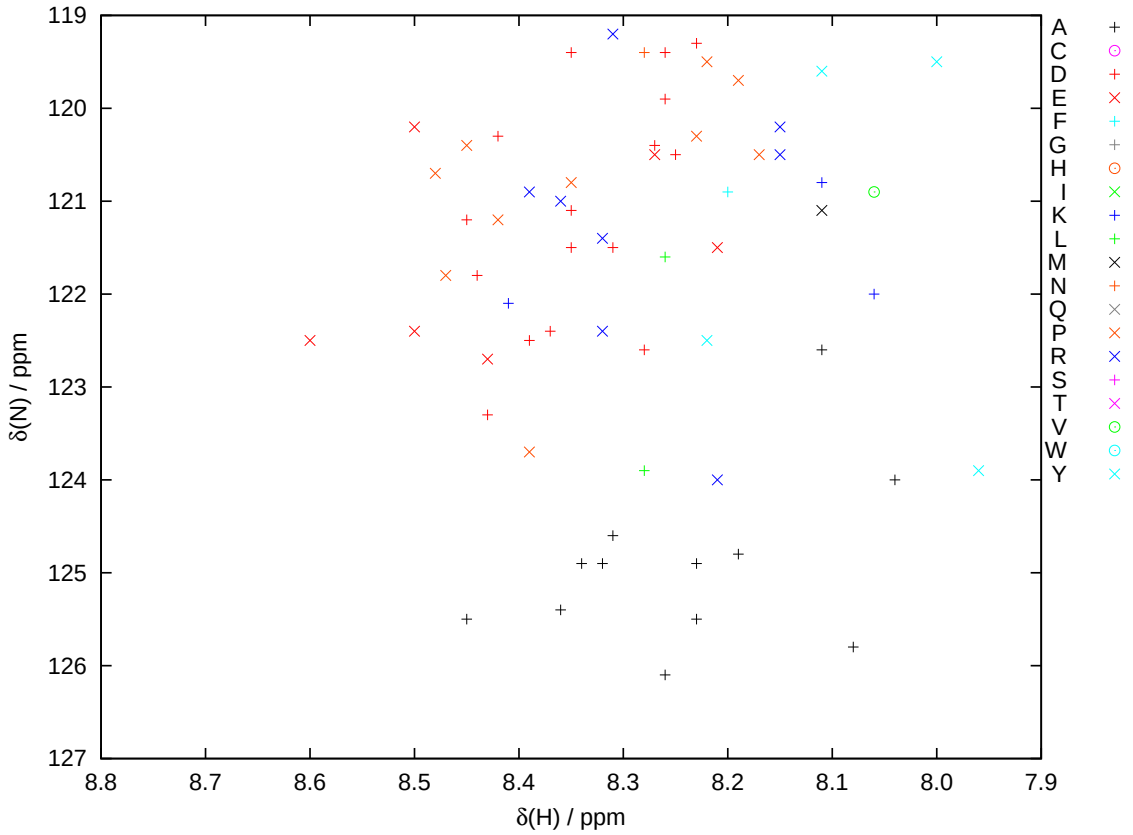

Figure S6: Correlation of ^{1}H and ^{15}N chemical shifts of a natively unfolded protein XC4149 from the plant pathogen *Xanthomonas campestris* pv. *campestris* (Ko-Hsin et al., BMRB Accession Number 6436). The amino acid types are distinguished by point size and color, as indicated at the right-hand side.

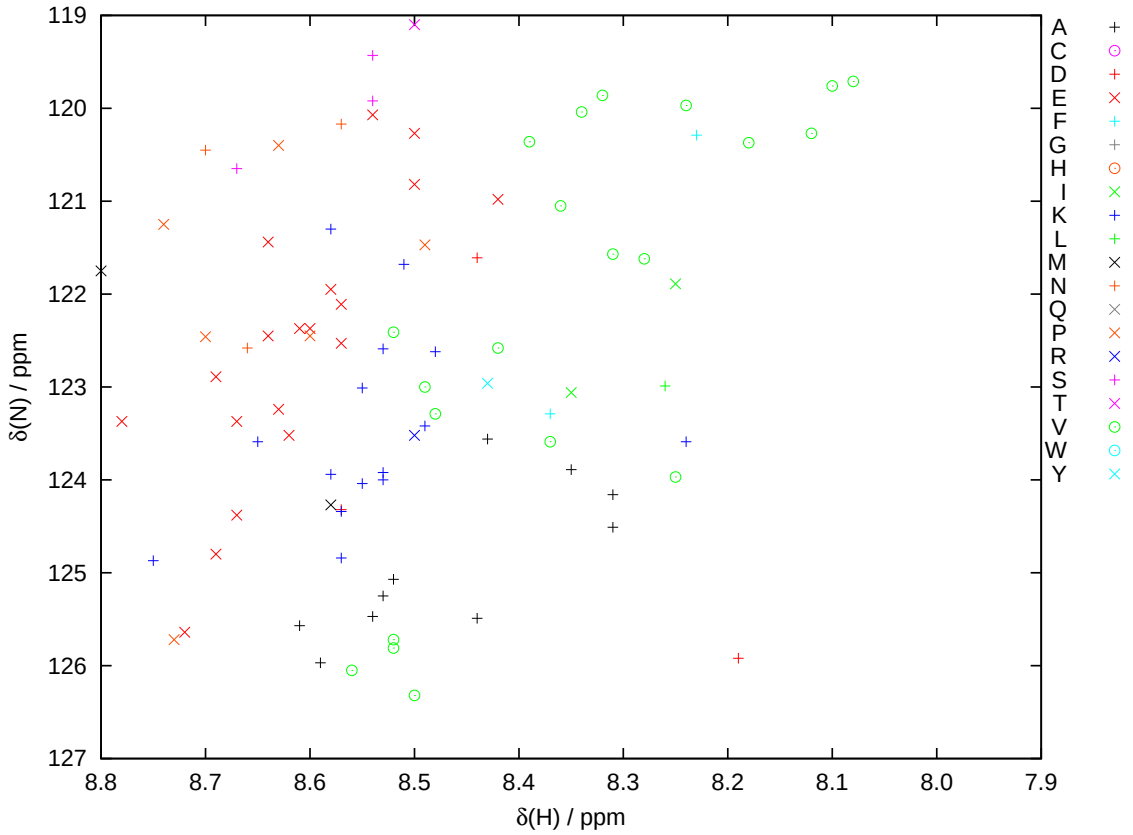

Figure S7: Correlation of ^{1}H and ^{15}N chemical shifts of a γ-synuclein (Marsh et al. (2006) *Protein Sci.* **15**, 2795–2804., BMRB Accession Number 7244). The amino acid types are distinguished by point size and color, as indicated at the right-hand side.

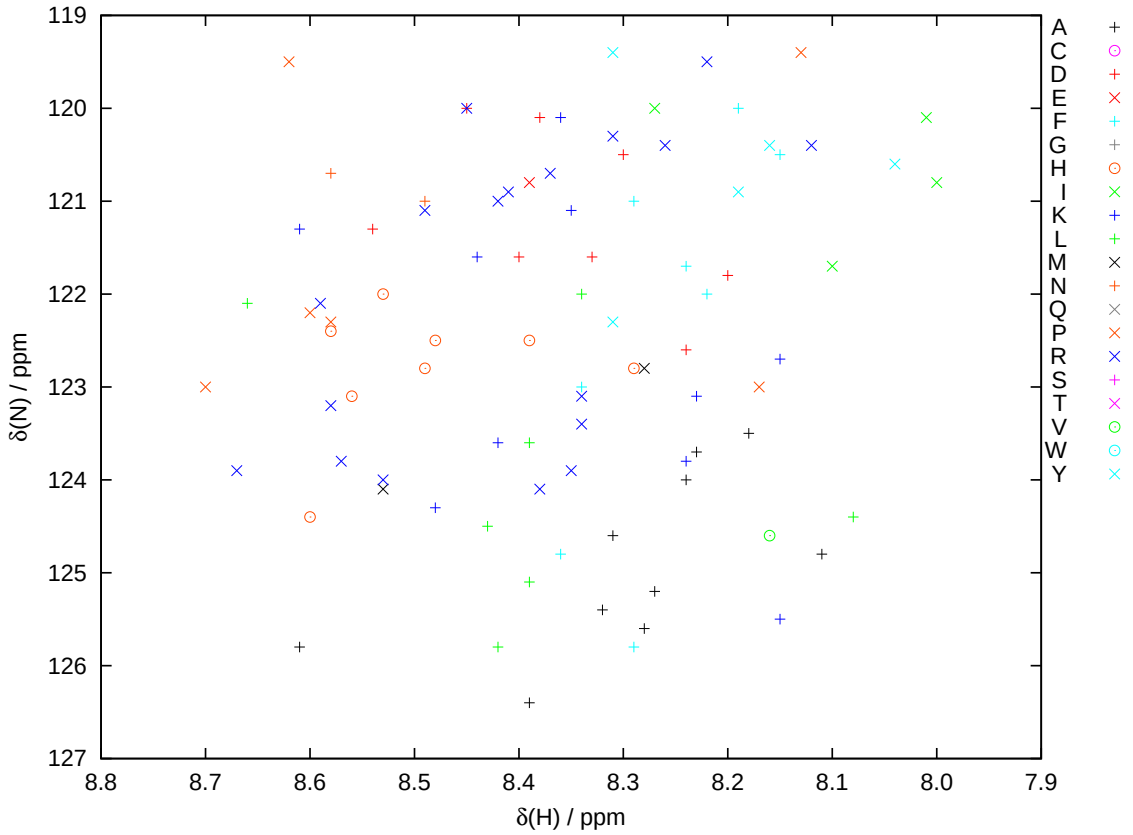

Figure S8: Correlation of ^{1}H and ^{15}N chemical shifts of a murine myelin basic protein (Libich et al. (2007) *Biomol. NMR Assign.* **1**, 61–63., BMRB Accession Number 15131). The amino acid types are distinguished by point size and color, as indicated at the right-hand side.

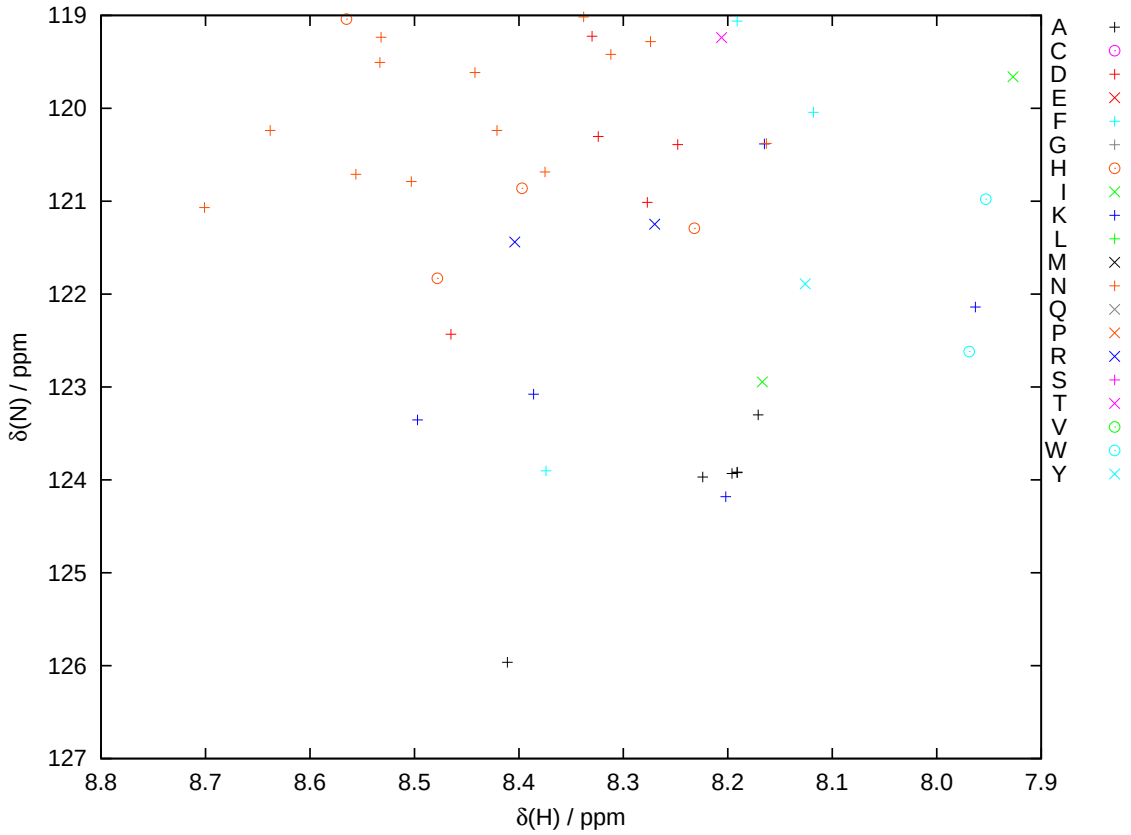

Figure S9: Correlation of ^{1}H and ^{15}N chemical shifts of intrinsically disordered translocation domain of colicin N (Hecht et al. (2008) *FEBS Lett.* **582**, 2673–2677., BMRB Accession Number 15506). The amino acid types are distinguished by point size and color, as indicated at the right-hand side.

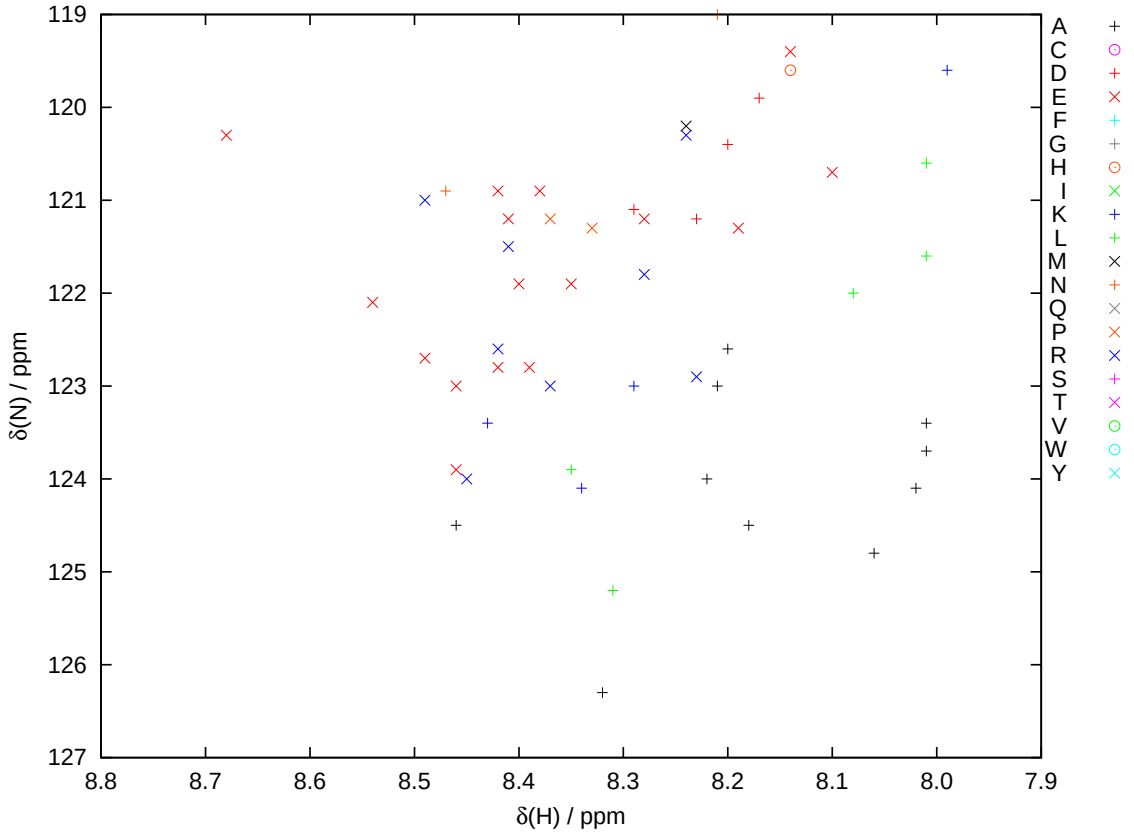

Figure S10: Correlation of ^{1}H and ^{15}N chemical shifts of the intrinsically disordered C-terminal region of *Homo sapiens* FCP1 in the unbound state (Showalter et al. (2009) *Biomol. NMR Assign.* **3**, 179-181., BMRB Accession Number 16296). The amino acid types are distinguished by point size and color, as indicated at the right-hand side.

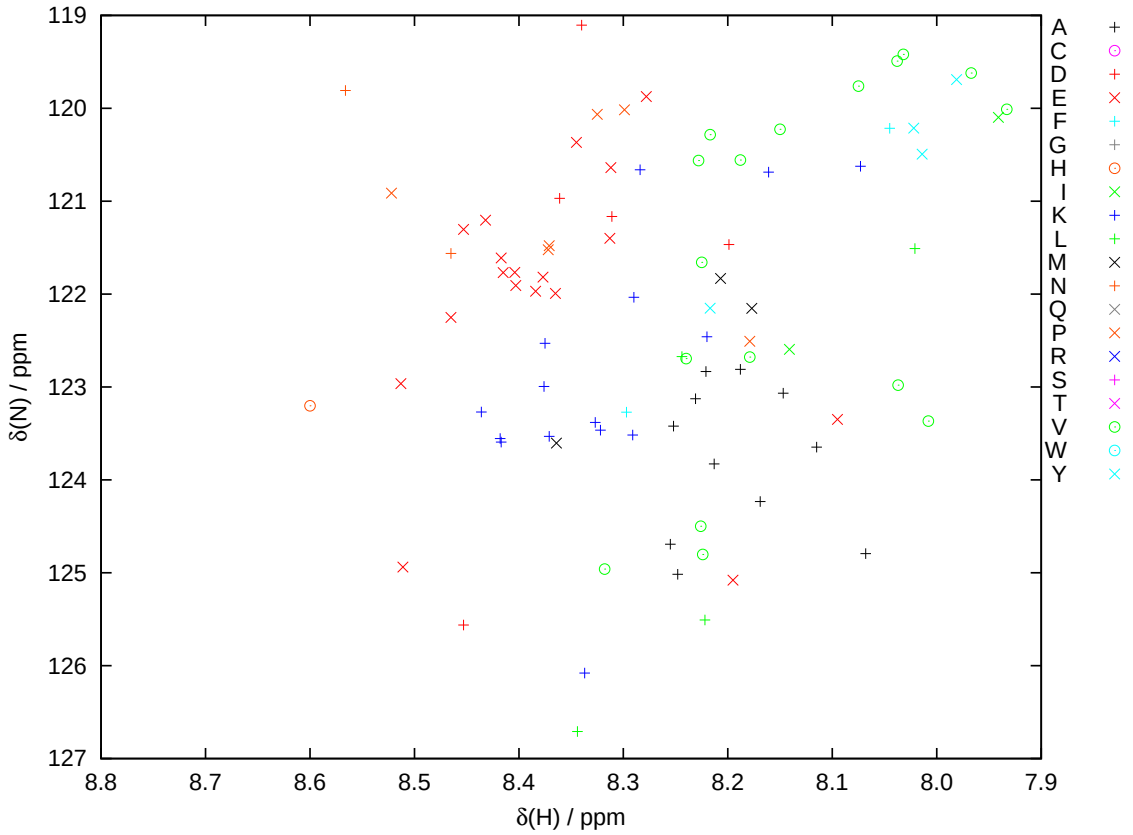

Figure S11: Correlation of ^{1}H and ^{15}N chemical shifts of disordered α-synuclein at pH 6.0 (Bodner et al. (2010) *Biochemistry* **49**, 862–871., BMRB Accession Number 16543). The amino acid types are distinguished by point size and color, as indicated at the right-hand side.

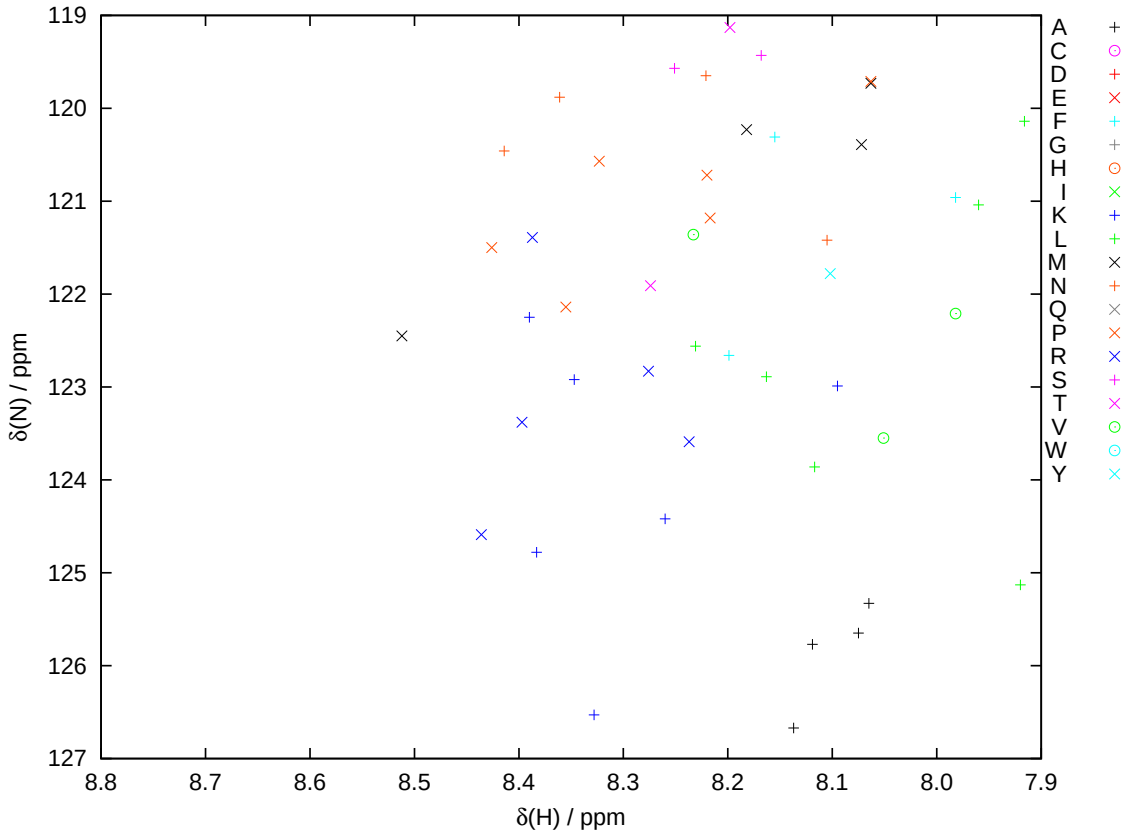

Figure S12: Correlation of ^{1}H and ^{15}N chemical shifts of intrinsically disordered CDK inhibitor Sic1 from yeast (Mittag et al. (2010) *Structure* **18**, 494–506., BMRB Accession Number 16657). The amino acid types are distinguished by point size and color, as indicated at the right-hand side.