STRUCTURE AND FUNCTION OF THE N-TERMINAL NUCLEOLIN BINDING DOMAIN OF NUCLEAR VALOCIN CONTAINING PROTEIN LIKE 2 (NVL2) HARBORING A NUCLEOLAR LOCALIZATION SIGNAL

Yoshie Fujiwara[1,2,3], Ken-ichiro Fujiwara[4,5], Natsuko Goda[1,3], Naoko Iwaya[1,3,6], Takeshi Tenno[1,2], Masahiro Shirakawa[6], Hidekazu Hiroaki[1,2,3,7]

[1]Division of Structural Biology, Graduate School of Medicine, Kobe University, 7-5-1 Kusunokicho, Chuo-ku, Kobe, Hyogo 650-0017, Japan

[2]Global-COE (Center of Excellence) Program for Integrative Membrane Biology, Kobe University, 7-5-1 Kusunokicho, Chuo, Kobe, Hyogo 650-0017, Japan

[3]Institute for Bioinformatics Research and Development (BIRD), Japan Science and Technology Corporation (JST), Japan

[4]Field of Supramolecular Biology, International Graduate School of Arts and Sciences, Yokohama City University, 1-7-29 Suehiro, Tsurumi, Yokohama, Kanagawa 230-0045, Japan

[5]Shionogi Research Laboratories, Shionogi & Co., Ltd., 5-12-4 Sagisu, Fukushima-ku, Osaka 553-0002, Japan

[6]Department of Molecular Engineering, Graduate School of Engineering, Kyoto University, Katsura, Kyoto 615-8510, Japan

Running title: Structure of an NVL2 N-terminal nucleolin binding domain

Supplemental Materials

Supplementary Experimental procedures

NMR Spectroscopy—Samples for NMR spectroscopy contained either ^{15}N- or ^{13}C/^{15}N-labeled NVL2^{UD} at a concentration of 0.6–1.2 mM in 5% D_2O-95% H_2O, 25 mM sodium phosphate (pH 6.4). Backbone and side chain assignments were obtained using combinations of ^{15}N-HSQC, ^{13}C-HSQC, HNCA, HNCO, HNCACB, CBCACONH, H(CCO)NH, C(CO)NH, and HCCH-TOCSY spectra recorded on 500 MHz, 600 MHz and 800 MHz NMR spectrometers (Bruker Biospin, Bruker Advance; Bruker Analytik GmbH, Germany) equipped with a cryomagnetic probe at 25 °C (1). Data were processed using the NMRPipe (2) and Sparky software (3). Inter-proton distances were obtained from either three-dimensional ^{13}C- or ^{15}N-edited NOESY with a 100 ms mixing time. Chemical shift assignments were deposited in the BioMagResBank under the accession code 11250.

Structure Analysis—Structures were calculated using the standard six-iteration cycle of the CYANA (version 2.0.7) program (4,5), followed by a simulated annealing refinement by CNS (version 1.2) (6). All NOE cross peaks were selected manually using Sparky. In total, 994 meaningful NOE upper distance restraints were obtained, including 236 with long-range information. Backbone torsion angle restraints of PHI and PSI in the regular secondary structure regions were derived using the results from the TALOS program (7), and were used during the

CYANA and CNS calculations. Starting from 200 initial structures, the 20 structures with the lowest energies were analyzed (Supplementary Table 1). The coordinates have been deposited in the Protein Data Bank (PDB) under the accession code 2rre. The chemical shift assignments have been deposited in the BioMagResBank under the accession code 11250 (http://www.bmrb.wisc.edu/).

Identification of NVL2^{UD}-associated proteins by mass spectrometry—Protein bands were excised from the gel and digested with trypsin. Peptide mass fingerprints of tryptic peptides were collected by LC-MS/MS using Q-TOF 2 (Micromass, Manchester, UK). Protein identifications were performed using the NCBInr database by the Mascot search engine (Matrix Science Inc., Boston, MA). A peptide mass tolerance of 0.2 Da and a fragment ion mass accuracy tolerance of 0.2 Da were used. Proteins with MOWSE protein scores >54 (significance threshold $p<0.05$) were considered statistically significant.

Western Blotting Analysis—HeLa and HEK293 cells were grown to the confluency in 150 mm dishes. Cells from 15 *HeLa* dishes and 10 HEK293 dishes were suspended in 5 ml of lysis buffer containing 25 mM Tris-HCl (pH 7.5), 100 mM NaCl, 2 mM EDTA, 0.1% (v/v) Triton X-100, 1 mM DTT, 1 mM PMSF and protease inhibitors. All subsequent procedures were described in the EXPERIMENTAL PROCEDURES.

NMR titration experiments—To examine the direct interaction between NVL2^{UD}(1-74) and nucleolin fragment R12, a series of ^{15}N-HSQC spectra with WATERGATE water suppression (8) were recorded at 30 °C, in 5% D_2O-95% H_2O containing 20 mM sodium phosphate buffer (pH 6.4). In the titration, 0, 1 and 3 molar equivalents of unlabeled nucleolin fragment R12 were added to 0.1 mM ^{15}N-labeled NVL2^{UD}(1-74).

Supplementary Figure S1

Ramachandran plot for the phi–psi values of the final 20 structures of NVL2^{UD}(1–74). This figure was produced using PROCHECK-NMR (9).

Supplementary Figure S2

Association of NVL2^{UD} with nucleolin. *A* & *B*. A Hela cell extract (A) or a HEK293 cell extract (B) was incubated with either GST (lanes "G"), GST-NVL2^{UD}(1-74) (lanes "74") or

GST-NVL2^{UD}(1-93) (lanes "93"). The purified proteins were separated by SDS-PAGE and immunoblotted using anti-nucleolin antibodies.

Supplementary Figure S3

Multiple alignment of human nucleolin RRM domains. Four sequences were aligned using clustalX (10) and manually adjusted. RNP1 and RNP2 indicate the positions of consensus sequences.

Supplementary Figure S4

The nucleolin-RNA complex is required to associate with NVL2^{UD}. The HSQC spectra of NVL2^{UD}(1-74), collected in the absence (black), or presence of 3 molar equivalents (red) of nucleolin fragment R12, indicate that NVL2^{UD} shows no specific interaction between NVL2^{UD} and purified (RNA-free) nucleolin R12.

References

1. Cavanagh, J., Fairbrother, WJ., Palmer, AG. III. , and Skelton, NJ. (2007) *San Diego, Academic Press.* 535-673
2. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J Biomol NMR* **6**, 277-293
3. Goddard, TD., and Kneller, DG. (2004) *University of California, San Francisco*
4. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) *J. Mol. Biol.* **319**, 209-227
5. Guntert, P. (2003) *Prog. Nuc. Magn. Reson. Spect.* **43**, 105-125
6. Brunger AT, , Adams PD, , Clore GM, , DeLano WL, , Gros P, , Grosse RWK, , Jiang JS, , Kuszewski J, , Nilges M, , Pannu NS, , Read RJ, , Rice LM, , Simonson T, , and Warren GL (1998) *Acta Crystallogr D Biol Crystallogr* **54**, 905-921
7. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J. Biomol. NMR* **13**, 289-302
8. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J Biomol NMR* **2**, 661-665
9. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR* **8**, 477-486
10. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res* **25**, 487648-82

Table S1.
Structural statistics for NVL2^{UD}.[a]

Total number of distance constrains	994
long range [| i-j | > 4]	236
middle range [1 < | i-j | ≤ 4]	376
short range [| i-j | ≤ 1]	382
Hydrogen bond constraints	20x2
Dihedral angle restraints	
ϕ, φ	46, 46
R.m.s. deviation from experimental constraints[b]	
Distance(Å)	$0.002\pm5\times10^{-4}$
Angle(°)	0.124±0.012
R.m.s. deviation from idealized covalent geometry	
Distance(Å)	$0.001\pm5\times10^{-5}$
Angle(°)	0.292±0.001
Impropers(°)	0.108±0.004
CNS energy terms (kcal/mol)	
E_{bond}	0.69±0.05
E_{angle}	29.8±0.23
E_{imp}	1.17±0.1
PROCHECK Ramachandran plot (residues 11-70) [c]	
Residues in most favored regions (%)	91.0
Residues in additionally allowed regions (%)	9.0
Residues in generously allowed regions (%)	0.0
Residues in disallowed regions (%)	0.0
R.m.s. deviation of mean structure derived from 20 calculated structures	
Back bone (residues 11-70) [c] (Å)	0.418
All heavy (residues 11-70) [c] (Å)	1.035

[a]These statistics comprise an ensemble of the 20 lowest-energy structures obtained from 200 starting structures.
[b]None of these structures exhibited distance violations > 0.5Å or dihedral angle violations > 5°.
[c]Residues 28 and 49-51 were excluded from analysis.

Table S2.

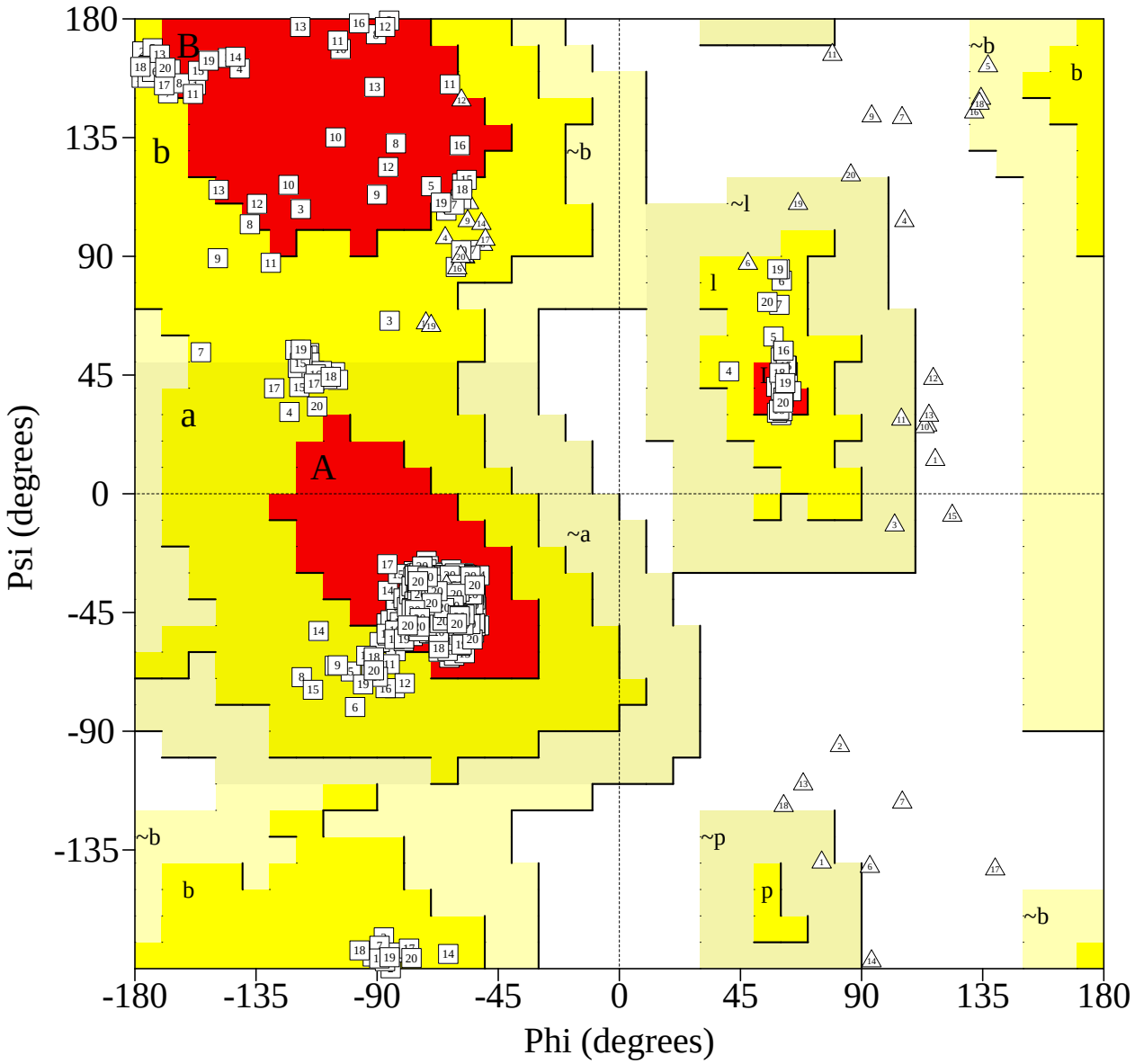
Gene name	peptide sequence	Score*	Rank
Nucleolin	EAMEDGEIDGNK	87	1
Progesterone receptor membrane component 2	GLGAGAGAGEESPATSLPR	78	1
Heterogeneous nuclear ribonucleoprotein D	IFVGGLSPDTPEEK	70	1
Heterogeneous nuclear ribonucleoprotein A3	EDTEEYNLR	68	1
Ribosomal protein L7a	AGVNTVTTLVENK	65	1
Prohibitin 2	IVQAEGEAEAAK	62	1
Vimentin	EEAENTLQSFR	62	1
Heterogeneous nuclear ribonucleoprotein C	SDVEAIFSK	55	1
Leucine rich repeat containing 59	LQQLPADFGR	44	1
Proteasome subunit, α type, 7	ALLEVVQSGGK	44	1
Proteasome subunit, α type, 5	ITSPLMEOSSIEK	42	1
Leucine rich repeat containing 59	DNPLDPVLAK	39	1
Ribosomal protein L7a	NFGIGQDIQPK	36	1
Proteasome subunit, α type, 5	AIGSASEGAQSSLQEVYHK	33	1
Heterogeneous nuclear ribonucleoprotein C	GFAFVQYVNER	25	1
Ribosomal protein L7a	VVNPLFEK	13	1
Proteasome subunit, α type, 7	ILNPEEIEK	19	4

*Score for the individual peptide

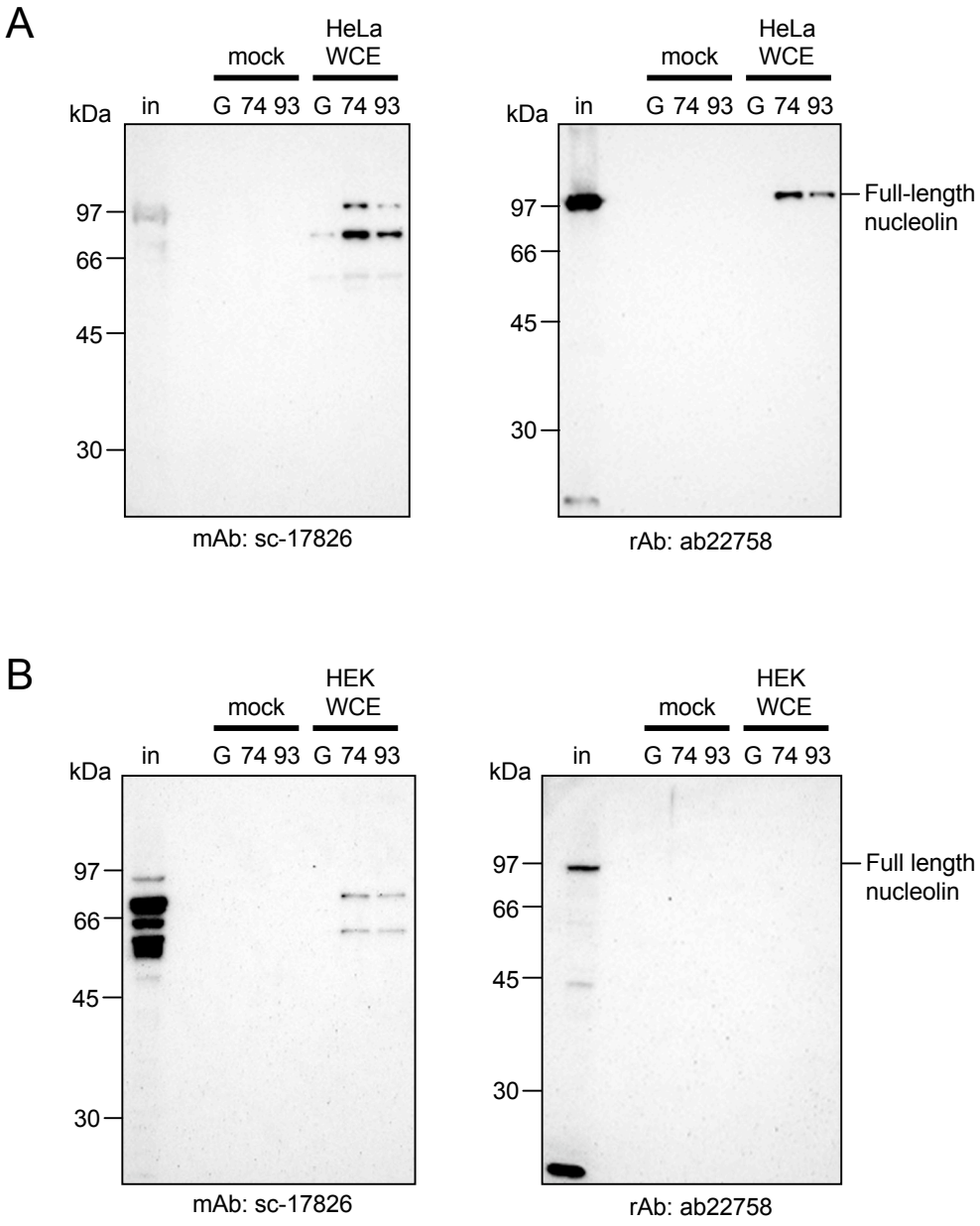
Table S3

Primers for cloning of human nucleolin

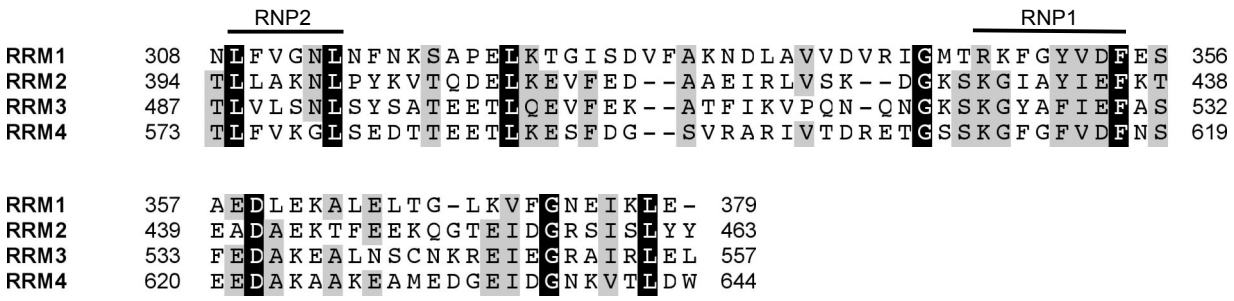
primer	sequence
NCL286-N	5’-AAACAGAAAGCAGCTCCTGAAGCC-3’
RRM1-C	5’-ACCTTATCGCTCTTTCTTACTGTCTTTTCC-3’
RRM2-N	5’-GATGCGAGAACACTTTTGG-3’
RRM2-C	5’-ACCTTAACCTTTCTCTCCAGTATAGTACAG-3’
RRM3-N	5’-ACCTTAGGGTCCTTGCAACTCCAG-3’
RRM3-C	5’-ACCTTAGGGTCCTTGCAACTCCAG-3’
RRM4-N	5’-AATGCCAGAAGCCAGCC-3’
RRM4-C	5’-ACCTTAGCCACCTTCACCCTTAGG-3’
GAR-N	5’-AAGGGTGAAGGTGGCTTC-3’
NCL-C	5’-ACCCTATTCAAACTTCGTCTTCTTTCCTTG-3’



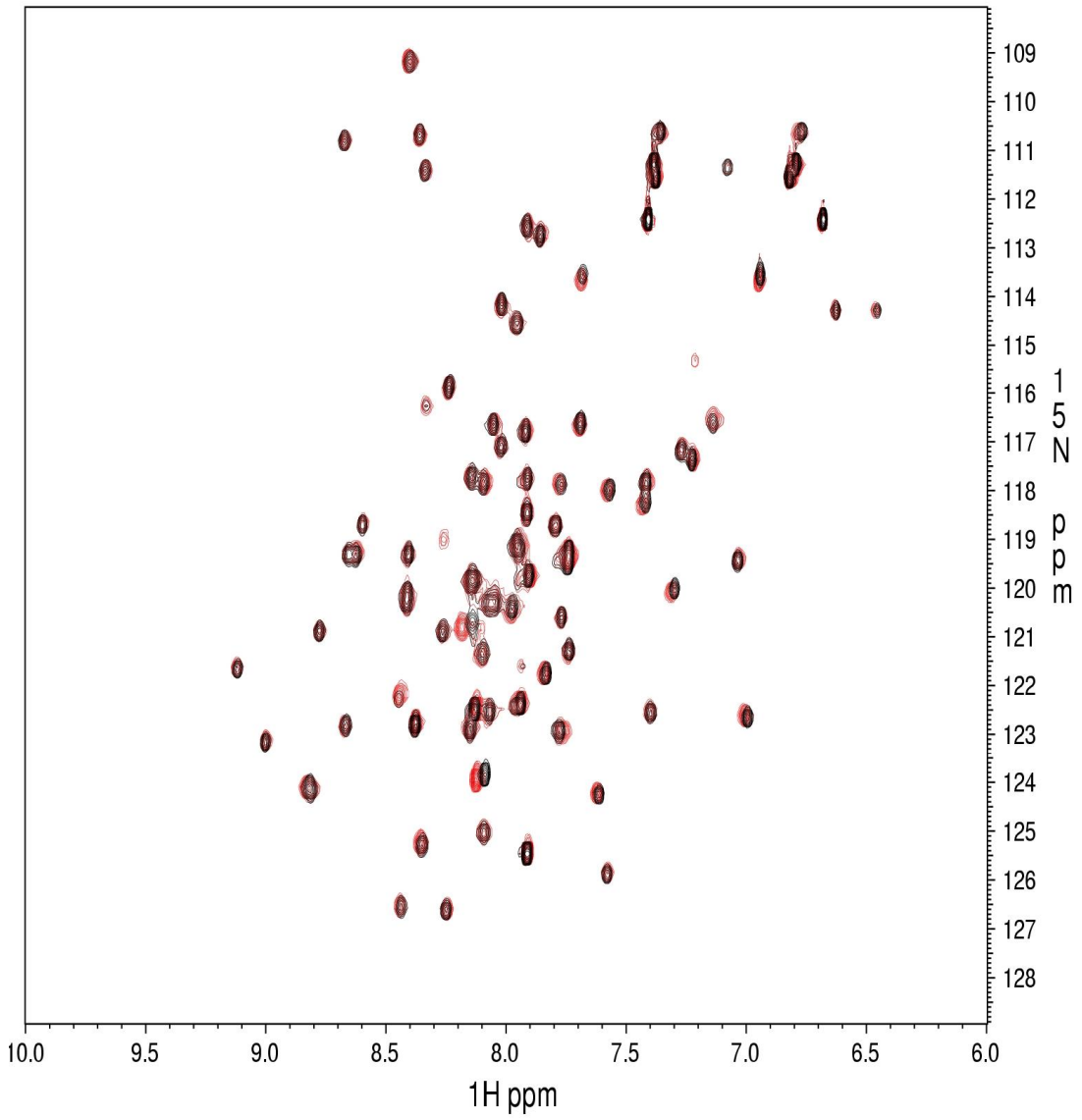
Supplementary Figure S1



Supplementary Figure S2



Supplementary Figure S3



Supplementary Figure S4