Journal of Biomolecular NMR **30:** 371–372, 2004.

Letter to the Editor: ^{1}H, ^{15}N and ^{13}C chemical shift assignments of RNA repeats binding protein – CUGBP1ab

Kyu-Yeon Jun[a], Youlin Xia[a], Xiaogang Han[a], Hua Zhang[a], Lubov Timchenko[b], Maurice S. Swanson[c] & Xiaolian Gao[a,*]

[a]*Department of Chemistry, University of Houston, Houston, TX 77004-5003, U.S.A.;* [b]*Department of Medicine, Baylor College of Medicine, One Baylor Plaza, Houston, TX, U.S.A.;* [c]*Department of Molecular Genetics and Microbiology, Centers for Gene Therapy and Mammalian Genetics, University of Florida College of Medicine, Gainesville, FL 32610, U.S.A.*

Received 2 June 2004; Accepted 27 July 2004

Key words: backbone assignments, CUG binding protein, neurodegenerative diseases, NMR solution structure, triplet repeats

Biological context

CUG-binding protein (CUGBP1) is an RNA-binding protein discovered in relation to myotonic dystrophy type 1 (DM1), a hereditary neurodegenerative disease (Timchenko et al., 1996). The onset of DM1 is associated with the expansion of an unstable $(CTG)_n$ repeat located within the 3′-untranslated region (3′-UTR) of the DMPK gene (Fu et al., 1992; Mahadevan et al., 1992). The protein was first identified due to its binding to the CUG triplet repeats (Timchenko et al., 1996). The investigation of CUGBP1 indicates that while the normal functions of protein may involve alternative splicing, translation and deadenylation, the aberrant activities of CUGBP1 affect several important cellular functions, such as overexpression of MEF2A and p21 to inhibit myogenesis, leading to DM1 syndrome manifested by muscle deficiency (Timchenko et al., 2004). CUGBP1 was found to be a signaling mediator in the epithelial growth factor receptor (EGFR) pathway affecting tissue growth and differentiation (Baldwin et al., 2004).

CUGBP1 contains three RNA-binding domains (RBDs) of about 70–80 residue per domain. RBD1 and RBD2 are adjacent to each other at the N-terminus and the RBD3 is close to the C-terminus, which is separated from RBD2 by a three residue transplicing variant region, a 20 residue Ser rich region, and a 100 residue linker domain of unknown function. It was suggested that RBD1 and RBD2 are necessary for specific binding to the CUG repeats (Timchenko et al., 1999), and thus CUGBP1ab containing RBD1 and RBD2 was constructed and studied. The binding of the protein to $(CUG)_8$ was confirmed by the gel-shift assays (data not shown). Additionally, recent report reveals that the UG dinucleotide repeats exhibit higher binding affinity to CUGBP1 than the CUG repeats does (Takahashi et al., 2000; Kino et al., 2004). We report herein the complete backbone and side chain assignments of CUGBP1ab as a first step to better understand the RNA binding properties associated with the protein. Since human genome is replete of repeating sequences of mostly unknown functions, the insights into repeating RNA-protein interactions may provide molecular basis for the biological role of repeating sequences in general.

*To whom the correspondence should be addressed. E-mail: xgao@uh.edu, Fax: 713-743-2709

Methods and experiments

The DNA sequence encoding resides 1-187 of CUGBP1 (gi:5729794) plus N-terminus His6 tag-thrombin cleavage site (21 aa) and a C-terminus tail (20 aa) was cloned into the NdeI/BamHI-site of the pET-15b vector (Novagen). The vector was transformed into *E. coli* BL21 (DE3) (Novagen) for overexpression of the protein. The cells were grown in 1 L M9 minimal medium, containing $^{15}NH_4Cl$ and ^{13}C-glucose (Cambridge Isotopes Labs) as the sole nitrogen and carbon source. When $OD_{600} = 0.8$ was reached, the protein expression was induced by adding

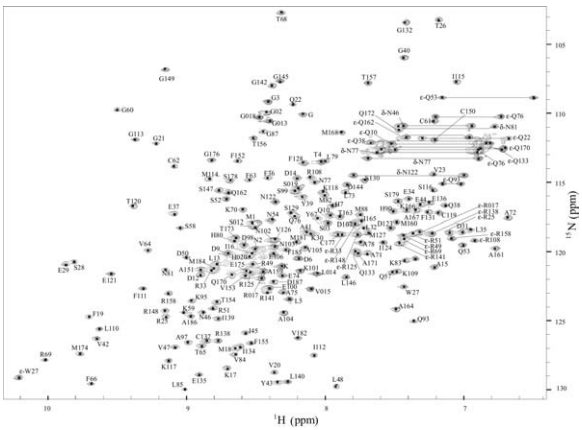

Figure 1. 2D ^{15}N-^{1}H HSQC of uniformly ^{15}N/^{13}C-labled CUGBP1ab. Assigned residues are marked with their residue name and residue number.

IPTG to a final concentration of 1 mM. After cultivating overnight at 30 °C, the cells were harvested. The protein was purified using a His-tag affinity column (Ni-NTA Superflow, Qiagen) and the purity was checked by SDS-PAGE. The final sample condition for NMR was ∼0.5 mM protein in 25 mM sodium phosphate, 50 mM NaCl, 0.25 mM NaN_3, and 0.125 mM EDTA at pH 5.8.

All the NMR spectra were recorded at 25 °C on a Bruker Avance 800 MHz NMR spectrometer with triple-resonance probe (^{1}H, ^{13}C, ^{15}N) including shielded z-gradient. Data processing and analysis were carried out as described elsewhere (Xia et al., 2004). Proton chemical shifts were calibrated with respect to water signal relative to DSS ($(CH_3)_3Si(CH_2)_3SO_3Na$); ^{15}N and ^{13}C chemical shifts were indirectly referenced to DSS. Sequence-specific backbone assignments were made from 3D HNCACB, CBCA(CO)NH, HBHANH, and HBHA(CO)NH. ^{1}H and ^{13}C side-chain assignments were performed with HCCH-COSY, HCCH-TOCSY, H(CCCO)NH and C(CCO)NH (mixing time: 16 ms). The NH_2 resonances from Asn and Gln side chains were assigned using 3D 2TS H-CN-H HSQC-NOESY-HSQC (mixing time: 120 ms; Xia et al., 2003). The NH_ϵ resonances of all Arg residues were identified from HNCACB and HBHANH. The $^{1}H_\delta$ and $^{1}H_\epsilon$ resonances of aromatic residues were assigned using 2D (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE.

Extent of assignments and data deposition

The ^{1}H-^{15}N HSQC spectrum of the ^{15}N/^{13}C-labeled CUGBP1ab is shown in Figure 1. All ^{1}H, ^{15}N and ^{13}C backbone and side-chain resonances of the protein have been assigned, except for the NHs of residues H89, I92, M94 and V183, the C_γ of Q22 and M88. These residues are mostly at the beginning or the end of the RBD domains. Additionally, the side-chain ^{13}C carbonyl resonances and partial aromatic ^{13}C and ^{1}H resonances were not assigned. The sequence-specific assignments have been deposited in the BioMagResBank database, accession number 6121 (www.bmrb.wisc.edu).

Acknowledgements

This research was supported by grants from NIH (GM49957) and the Robert A. Welch Foundation (E-1027). Keck/IMD NMR Center (www.chem.uh.edu/imd/nmr) is funded in part by the W. M. Keck Foundation.

References

Baldwin, B.R., Timchenko, N.A. and Zahnow, C.A. (2004) *Mol. Cell. Biol.*, **24**, 3682–3691.

Fu, Y.H., Pizzuti, A., Fenwick, R.G., King, J., Rajnarayan, S., Dunne, P.W., Dubel, J., Nasser, G.A., Ahizawa, T., deJong, P., Wieringa, B., Korneluk, R., Perryman, M.B., Epstein, H.F. and Caskey, C.T. (1992) *Science*, **255**, 1256–1258.

Kino, Y., Mori, D., Oma, Y., Takeshita, Y., Sasagawa, N. and Ishiura, S. (2004) *Hum. Mol. Genet.*, **13**, 495–507.

Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, S., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, J., O'Hoy, K., Leblond, S., Earle-M., J., deJong, P.J., Wieringa, B. and Korneluk, R.G. (1992) *Science*, **255**, 1253–1255.

Takahashi, N., Sasagawa, N., Suzuki, K. and Ishiura, S. (2000) *Biochem. Biophys. Res. Commun.*, **277**, 115–121.

Timchenko, L.T., Miller, J.W., Timchenko, N.A., DeVore, D.R., Datar, K.V., Lin, L. Roberts, R., Caskey, C.T. and Swanson, M.S. (1996) *Nucl. Acids Res.*, **24**, 4407–4414.

Timchenko, N.A., Patel, R., Iakova, P., Cai, Z.J., Quan, L. and Timchenko, L.T. (2004) *J. Biol. Chem.*, **279**, 13129–13139.

Timchenko, N.A., Welm, A.L., Lu, X. and Timchenko, L.T. (1999) *Nucl. Acids Res.*, **27**, 4517–4525.

Xia, Y., Yee, A., Arrowsmith, C.H. and Gao, X. (2003) *J. Biomol. NMR*, **27**, 193–203.

Xia, Y., Zhu, G., Veeraraghavan, S. and Gao X. (2004) *J. Biomol. NMR*, **29**, 467–476.