Journal of Biomolecular NMR **30:** 357–358, 2004.

Letter to the Editor: ^{1}H, ^{13}C and ^{15}N backbone assignment of a 32 kDa hypothetical protein from *Arabidopsis thaliana*, At3g16450.1

Nozomi Sugimori[a], Takuya Torizawa[b], David J. Aceti[c], Sandy Thao[c], John L. Markley[c] & Masatsune Kainosho*[a,b]

[a]*Graduate School of Science, Tokyo Metropolitan University, 1-1 Minami-ohsawa, Hachioji, 192-0397, Japan;* [b]*CREST, Japan Science and Technology Agency (JST), 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan;* [c]*Center for Eukaryotic Structural Genomics, Department of Biochemistry, 171A Biochemistry Addition, University of Wisconsin, 433 Babcock Drive, Madison, WI 53706-1344, USA*

Received 22 March 2004; Accepted 7 July 2004

Key words: Arabidopsis thaliana, hypothetical protein, NMR assignments, structural genomics

Biological context

The flowering brassica *Arabidopsis thaliana* is important as a model for identifying plant genes and determining their functions. Its many advantages include a relatively small genome of about 130 Mb and lack of repetitious DNA (Meinke et al., 1998). Its complete genome sequence has been released, and its functional and structural properties are being addressed (The Arabidopsis Genome Initiative, 2000). At3g16450.1 is a 32 kDa, 299-residue hypothetical protein from *A. thaliana*. Homology search methods, such a Reverse PSI-BLAST (Marchler et al., 2002), reveal similarity between the At3g16450.1 sequence and those of certain lectins. The proteins showing the highest homology with the N- and C-terminal halves of the At3g16450.1 sequence are artocarpin (30% identity) specific to mannose (Pratap et al., 2002), and agglutinin (34% identity) specific to a tumor antigen, Galβ1,3GalNAc (Lee et al., 1998), respectively. Thus, At3g16450.1 is expected to be a bifunctional lectin. Although three-dimensional structures of proteins homologous to each of the two domains of At3g16450.1 have been determined by X-ray crystallography, there have been no reports of NMR resonance assignment from this family. Therefore, the assigned chemical shifts of of At3g16450.1 should be useful, not only for future solution structural determinations, but also for characterizing the sugar binding features of the family. Here, we report assignments for the backbone ^{1}H, ^{13}C and ^{15}N and side chain $^{13}C^{\beta}$ resonances of the protein doubly labeled with ^{13}C and ^{15}N.

*To whom correspondence should be addressed. E-mail: kainosho@nmr.chem.metro-u.ac.jp

Methods and experiments

Primers were designed and used to clone the At3g16450.1 gene by PCR from a cDNA library derived from the T87 tissue culture of *Arabidopsis thaliana* ecotype Columbia as described previously (Aceti et al., 2003). NdeI and BamHI restriction sites were used in subcloning the gene into the pET15b expression vector (Novagen) containing an N-terminal hexa-His purification tag. The recombinant protein was overexpressed at 30 °C for 6 h in BL21 (DE3) *E. coli* cells (Invitrogen) grown in a minimum medium containing 1 g/l $^{15}NH_4Cl$ (Shoko) and 2 g/l $^{13}C_6$-glucose (Shoko) and supplemented with 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 2.4 mg/l thiamine, 3.5 mg/l $FeCl_3$ and 1.2 mg/l Biotin. Cell pellets were lysed in 50 mM Tris-HCl pH 8.0 (Buffer A), loaded on Hi-Trap DEAE FF (Amersham Biosciences) and eluted with a linear 0-1 M NaCl gradient. The eluted protein was further purified by Ni-NTA agarose (Qiagen) according to methods in provided by the manufacturer. After dialysis against Buffer B (Buffer A containing 150 mM NaCl and 2.5 mM $CaCl_2$), the recombinant protein was digested with thrombin protease (Amersham Biosciences) (100 units/mg of the protein) at room temperature for 2 h to cleave the peptide containing the hexa-His tag from At3g16450.1. Consequently the protein

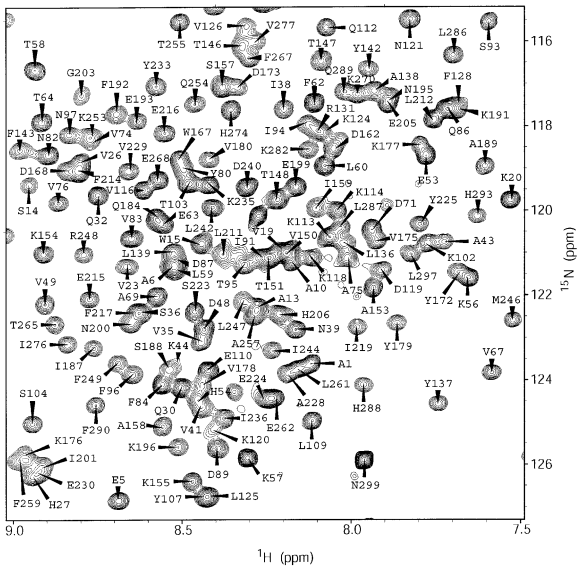

Figure 1. ^{1}H-^{15}N HSQC spectrum of At3g16450.1 acquired at 27.5 °C, pH 6.8. This is a zoomed region of the most crowded area of the spectrum. Assignments are indicated alongside the corresponding signals with the one-letter amino acid code and residue number. Signals without a residue number were not assigned.

product contained an additional tetrapeptide (Gly-Ser-His-Met) at the N-terminus. The reaction solution was applied to three successive columns equilibrated with Buffer A containing 150 mM NaCl (Buffer C): Benzamidine Sepharose 6B (Amersham Biosciences), Chelex 100 Resin (Bio-Rad), and Ni-NTA agarose. Each column was washed with Buffer C, and the protein was collected in the flow-through. The purified protein solution was both concentrated and buffer-exchanged by centrifugation in VIVASPIN 20 (molecular weight cutoff 10,000; Sartorius). The NMR sample containing 0.76 mM At3g16450.1 in 20 mM bis-Tris-d_{19} (Cambridge Isotope Laboratories), 100 mM KCl, 10% D_2O, pH 6.8 was recorded at 27.5 °C on a Bruker DRX600 NMR spectrometer equipped with 5-mm TXI CryoProbe. All spectra were processed with XWINNMR version 3.5 (Bruker) and analyzed with SPARKY version 3.106 (Goddard and Kneller, University of California, San Francisco). Sequence specific resonance assignments were carried out by standard methods (reviewed in: Grzesiek and Bax, 1992; Grzesoel et al., 1993; Clore and Gronenborn, 1994). Backbone signals were assigned by analyzing ^{1}H-^{15}N HSQC, HNCO, HN(CA)CO, CT-HNCA, CT-HN(CO)CA, HNCACB, HN(CO)CACB, HBHANH, and HBHA(CO)NH data sets. ^{1}H chemical shifts were referenced to external DSS, and ^{15}N and ^{13}C shifts were indirectly referenced (Wishart, et al., 1995; Markley et al., 1998).

Extent of assignments and data deposition

Figure 1 shows ^{1}H-^{15}N HSQC spectrum of At3g16450.1. The backbone and sidechain beta resonances were assigned to the following extents: HN except for Pro (96%), N except for Pro (96%), $^{13}C'$ (93%), $^{13}C^{\alpha}$ (98%), $^{13}C^{\beta}$ (92%) and $^{1}H^{\alpha}$ (89%). The assigned ^{1}H, ^{13}C and ^{15}N chemical shifts for At3g16450 have been deposited in the BioMagResBank database (http://www.bmrb.wisc.edu) under accession number 6140.

Acknowledgements

This work was supported by CREST/JST and NIH grant 1 P50 GM64598.

References

Aceti, D.J., Blommel, P.G., Endo, Y., Fox, B.G., Frederick, R.O., Hegeman, A.D., Jeon, W.B., Kimball, T.L., Lee, J.M., Newman, C.S., Peterson, F.C., Sawasaki, T., Seder, K.D., Sussman, M.R., Ulrich, E.L., Wrobel, R.L., Thao, S., Vinarov, D.A., Volkman, B.F. and Zhao, Q. (2003) In *Biopolymers*, Vol. 8, Steinbüchel, A. (Ed.), Wiley-VCH, Weinheim, pp. 469–496.

Clore, G.M. and Gronenborn, A.M. (1994) In *Nuclear Magn. Reson., Pt C*, Vol. 239, pp. 349–363.

Grzesiek, S. and Bax, A. (1992) *J. Magn. Reson.*, **99**, 201–207.

Grzesiek, S., Anglister, J. and Bax, A. (1993) *J. Magn. Reson. B*, **101**, 114–119.

Lee, X., Thompson, A., Zhang, Z., Toh-that, H., Biesterfeldt, J., Ogata, C., Xu, L., Johnston, R.A.Z. and Young, N.M. (1998) *J. Biol. Chem.*, **273**, 6312–6318.

Marchler-Bauer A., Panchenko, A.R., Shoemaker, B.A., Thiessen, P.A., Geer, L.Y., Bryant S.H. (2002) *Nucl. Acids Res*, **30**, 281–283.

Meinke, D.W., Cherry, J.M., Dean, C., Rounsley, S.D., Koornneef, M. (1998) *Science* **282**, 662–682.

Markley, J.L., Bax, A., Arata, Y., Hilbers, C.W., Kaptein, R., Sykes, B.D., Wright, P.E. and Wüthrich, K. (1998) *Pure Appl. Chem.*, **70**, 117–142.

Pratap, J.V., Jeyaprakash, A.A., Rani, P.G., Sekar, K., Surolia, A. and Vijayan, M. (2002) *J. Mol. Biol.*, **317**, 237–247.

The Arabidopsis Genome Initiative (2000) *Nature*, **408**, 796–815.