Journal of Biomolecular NMR **30:** 377–378, 2004.

Letter to the Editor: ^{1}H, ^{13}C, and ^{15}N chemical shift assignment of the C-terminal 15 kDa domain of a novel galactose-binding protein from the earthworm *Lumbricus terrestris*

Hikaru Hemmi[a,*], Atsushi Kuno[b], Shigeyasu Ito[c], Ryuichiro Suzuki[c], Satoshi Kaneko[a], Tsunemi Hasegawa[c], Jun Hirabayashi[b] & Ken-ichi Kasai[d]

[a]*National Food Research Institute, Tsukuba, Ibaraki 305-8642, Japan;* [b]*Research Center for Glycoscience, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8566, Japan;* [c]*Department of Material and Biological Chemistry, Yamagata University, Yamagata, Yamagata 990-8560, Japan;* [d]*Department of Biological Chemistry, Teikyo University, Sagamiko, Kanagawa 199-0195, Japan*

Received 21 June 2004; Accepted 20 July 2004

Key words: earthworm, galactose-binding lectin, NMR assignment, R-type lectin family

Biological context

A novel 29-kDa lectin (EW29) has been isolated from the earthworm *Lumbricus terrestris* by affinity chromatography on asialofetuin-agarose in the screening of galectin-like proteins (Hirabayashi et al., 1998). This lectin consists of two homologous domains (14,500 Da) showing 27% identity with each other and it has multiple short conserved motifs, 'Gly-X-X-X-Gln-X-Trp', in the sequence. This short motif has been found in many carbohydrate-recognition proteins from various organisms such as plant lectin ricin B-chain and *Streptomyces lividans* xylanase A. Carbohydrate-recognition proteins having the short conserved motif form the R-type lectin family. Although EW29 was prepared essentially by the same strategy as that used for galectin, this lectin appears to be a member of the R-type lectin family. EW29 has hemagglutinating activity differing from other tandem repeat-type proteins in the R-type lectin family such as ricin, abrin, and *Sambucus sieboldina* agglutinin, however. Based on structural features, this type of lectin generally contains one sugar-binding site per domain, suggesting that the truncated mutant comprising a single domain may have no hemagglutinating activity (Rutenber & Robertus, 1991; Tahirov et al., 1995; Kaku et al., 1996). The C-terminal domain of EW29 binds to asialofetuin-agarose as strongly as the whole protein (Hirabayashi et al., 1998) and retains its hemagglutinating activity 10-fold lower than the whole protein, whereas the N-terminal domain completely reduces its hemagglutinating activity (Hirabayashi, unpublished results). These results indicate that the C-terminal domain of EW29 has more than one sugar-binding site, but sugar-binding sites in the C-terminal domain have not been identified yet. R-type lectins are reported to have physiological functions such as enzyme targeting and glycoprotein hormone turnover (reviewed in Dodd and Drickamer, 2001). The physiological function of EW29, however, remains unknown.

Here we report chemical shift assignments for the C-terminal domain of EW29 from the earthworm *Lumbricus terrestris*. The solution structure of the C-terminal domain of EW29 arising from further analysis of these chemical shift assignments is expected to provide a useful comparison to the tertiary structures of other proteins in the R-type lectin family. The chemical shift mapping perturbation of the C-terminal domain of EW29 with sugar by ^{1}H-^{15}N HSQC spectroscopy is also expected to identify sugar-binding sites in the C-terminal domain of EW29. This work thus provides the basis for more detailed study of the interaction between the C-terminal domain of EW29 and the carbohydrate required for carbohydrate binding specificity.

Materials and experiments

The expression vector for truncated EW29 lectin composed of the C-terminal domain (C-half) was constructed as described elsewhere (Hirabayashi et al., 1998). The expression vector was transformed into *Escherichia coli* cells, Epicurian Coli® BL21-CodonPlus™ Competent Cell (Stratagene, CA, U.S.A.). Uniformly ^{15}N and ^{13}C double labeled C-half was expressed in doubly labeled C.H.L. medium (Chlorella Co., Japan) containing ampicillin (50 μg/ml), and the soluble pro-

*To whom correspondence should be addressed. E-mail: hemmi@nfri.affrc.go.jp

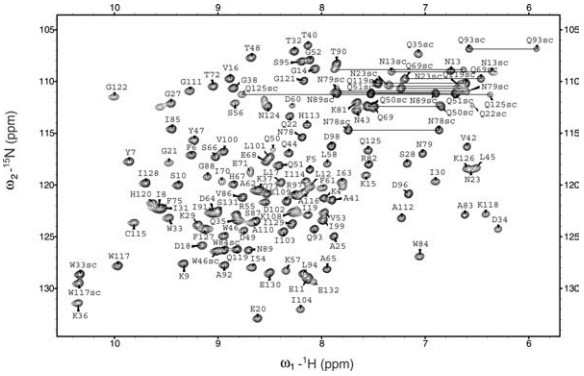

Figure 1. A 600 MHz 2D ^{1}H-^{15}N HSQC spectrum of the 0.9 mM C-terminal domain of EW29 at pH 6.1 and 298K. Cross-peaks are labeled upon the basis of an analysis of through-bond connectivities. Side chains of NH_2 resonances of asparagines and glutamines are connected by horizontal lines and are marked by 'sc'. Side chains of NH resonances of tryptophane are also marked by 'sc'.

tein was purified as described elsewhere (Hirabayashi et al., 1998; Ito et al., 2004).

NMR samples contained the 0.9 mM C-terminal domain of EW29 in 50 mM phosphate buffer (pH 6.1, 90:10 v/v H_2O/D_2O) and a protease inhibitor cocktail (Sigma Chemical Co., MO, USA). NMR experiments were recorded at 298 K with a Bruker Avance-600 spectrometer. Sequence-specific assignments of the polypeptide backbone were made from ^{1}H-^{15}N HSQC, HNCA, HNCO, HN(CO)CA, HNCACO, HACACONH, CBCACONH, HBHA(CBCA)NH and HNCACB spectra, while assignments of the side chain resonances were made from CCONH, HCCONH, HCCH-COSY, and HCCH-TOCSY spectra. Aromatic side chain resonances were assigned from 2D ^{1}H-NOESY and TOCSY, CT-^{13}C-HSQC, ^{13}C-edited NOESY, and TOCSY-CT-HSQC spectra recorded on natural abundance and [*U*-^{13}C, *U*-^{15}N]-labeled samples in the aromatic carbon region.

NMR data were processed using Felix2000 and analyzed using Sparky (T.D. Goddard and D.G. Kneller, SPARKY 3, University of California, San Francisco, CA, USA). All ^{1}H dimensions were referenced to internal 4,4-dimethyl-4-silapentane-1-sulfate (DSS), and ^{13}C and ^{15}N were indirectly referenced to DSS (Wishart et al., 1995).

Extent of assignments and data deposition

All ^{1}H and ^{15}N backbone resonances were assigned except for the first two residues of the C-terminal domain of EW29, and the residues N59, D76, and K105-D107. Figure 1 shows the ^{1}H-^{15}N HSQC spectrum for the C-terminal domain of EW29, and reveals 117 assigned ^{1}H-^{15}N backbone cross-peaks of the C-terminal domain of EW29 (95% complete) and 4 assigned ^{1}H-^{15}N side-chain cross-peaks of tryptophan residues in the C-terminal domain of EW29.

Of the 650 backbone resonance signals expected from the C-terminal domain of EW29, a total of 611 were observed and assigned. A number of backbone atoms were not observed, including (1) N, NH, Hα, Cα, and carbonyl carbon signals from residues K2, D76, K105, and S106; (2) Hα, Cα, and carbonyl carbon signals from M1 and I70; (3) N and NH signals from residues N59 and D107; and (4) N signal from all proline residues. Resonance assignments were also made for 875 of the 971 expected side-chain atoms including aromatic ring atoms (90% complete).

A secondary structure prediction was made based on ^{1}Hα, ^{13}Cα, ^{13}Cβ, ^{13}CO, and ^{15}N chemical shifts using CSI (Wishart et al., 1994) and TALOS (Cornilescu et al., 1999). These predictions suggest the presence of 10–12 β-strands, and the position of these secondary structure elements is in good agreement with predictions based on the amino acid sequence using PSIPRED (McGuffin et al., 2000).

^{1}H, ^{13}C, and ^{15}N chemical shifts for the C-terminal domain of EW29 have been deposited in the BioMagResBank database (http://www.bmrb.wisc.edu) under the accession number BMRB-6226.

Acknowledgement

This work was supported in part by the Program for Promotion of Basic Research Activities for Innovative Biosciences, Japan.

References

Cornilescu, G., Delaglio, F. and Bax, A. (1999) *J. Biomol. NMR*, **13**, 289–302.

Dodd, R.B. and Drickamer, K. (2001) *Glycobiology*, **11**, 71R–79R.

Hirabayashi, J., Dutta, S.K. and Kasai, K. (1998) *J. Biol. Chem.*, **273**, 14450–14460.

Ito, S., Kuno, A., Suzuki, R., Kaneko, S., Kawabata, Y., Kusakabe, I. and Hasegawa, T. (2004) *J. Biotechnol.*, **110**, 137–142.

Kaku, H., Tanaka, Y., Tazaki, K., Minami, E., Mizuno, H. and Shibuya, N. (1996) *J. Biol. Chem.*, **271**, 1480–1485.

McGuffin, L.J., Bryson, K. and Jones, D.T. (2000) *Bioinformatics*, **16**, 404–405.

Rutenber, E. and Robertus, J.D. (1991) *Proteins*, **10**, 260–269.

Tahirov, T.H., Lu, T.-H., Liaw, Y.-C., Chen, Y.-L. and Lin, J.-Y. (1995) *J. Mol. Biol.*, **250**, 354–367.

Wishart, D.S. and Sykes, B.D. (1994) *J. Biomol. NMR*, **4**, 171–180.

Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995) *J. Biomol. NMR*, **6**, 135–140.