Journal of Biomolecular NMR, **26:** 277–278, 2003.
KLUWER/ESCOM
© 2003 *Kluwer Academic Publishers. Printed in the Netherlands.*

Letter to the Editor: ^{1}H, ^{13}C and ^{15}N backbone resonance assignments of the N-terminal domain of *Drosophila* GCM protein

Masato Shimizu[a,b], Hidekazu Hiroaki[a,c,*], Daisuke Kohda[a,d], Eugene Hayato Morita[e], Shigeki Hotta[f] & Kosuke Morikawa[a]

[a]*Biomolecular Engineering Research Institute, 6-2-3 Furue-dai Suita Osaka, Japan 560-0874;* [b]*Venture Business Laboratory, Ehime University, 3-5 Bunkyo-cho, Matsuyama Ehime, Japan 790-8577;* [c]*Division of Molecular Biophysics, Graduate School of Integrated Sciences, Yokohama City University, 1-7-29, Suehiro-cho, Tsurumi-ku, Yokohama Kanagawa, Japan 230-0045;* [d]*Division of Structural Biology, Medical Institute of Bioregulation, Kyushu University 3-1-1 Maidashi, Higashi-ku, Fukuoka, Japan 812-8582;* [e]*Center for Gene Research, Ehime University, 3-5-7 Tarumi, Matsuyama Ehime, Japan 790-8566;* [f]*National Institute of Genetics, Mishima Shizuoka, Japan 411-8540*

Received 1 January 2003; Accepted 11 March 2003

Key words: DNA binding domain, glial cell missing, gliogenesis, transcription factor, zinc binding domain

Biological context

Glial Cell Missing (*gcm*) is a family of novel transcriptional factors which is involved in the early determination events of neural precursor cells in *Drosophila melanogaster* (Hosoya et al., 1995; Jones et al., 1995). *gcm*-deficient flies produce an increased number of neural cells at the expense of their glial counter part, whereas *gcm*-overexpressing flies have increased number of glial cells at the expense of the neurons. Therefore, *gcm* has been postulated to function as a binary genetic switch to determine the cell fates of neural precursor between glial and neuronal cell in *Drosophila*.

Several GCM-like proteins have also been identified in human, mouse and rat. Functional analysis of GCM-like proteins revealed that GCM is a family of transcriptional factors (Akiyama et al., 1996), which bind to the specific DNA sequence 5′-ATGCGGGT-3′ and its complementary sequence (Akiyama et al., 1996; Tuerk et al., 2000). Among the known seven (fly GCM1/2, mouse GCMa/b, human GCMa/b and rat GCMa) protein sequences, a well-conserved domain of approximately 150 amino acids was found at the N-terminus of the molecule. This domain was named 'GCM-box', which is indispensable for DNA binding (review: Wegner and Riethmacher, 2001). The GCM-box did not show any sequence homology to known proteins, whereas characteristic seven cysteine residues are critically conserved. Based on the analogy to other DNA binding zinc finger proteins, the GCM-box has also been thought to be a zinc-binding domain. However, the sequence signature of the possible zinc binding residues does not match any known Zn^{2+}-fingers or Zn^{2+}-clusters. Recently, Cohen et al. (2002) showed that GCM-domain from mouse GCMa bound to two zinc ions based on EXAFS studies, while the coordinating residues remained unclear (Cohen et al., 2002). In contrast, *in vitro* study of mouse GCMa ruled out the involvement of zinc coordination (Schreiber et al., 1998). In efforts to clarify the controversial situation of metal binding property of the GCM-box, we have undertaken NMR analysis of N-terminal DNA binding domain of *Drosophila* GCM protein. Here, we report the nearly complete sequence specific assignment of backbone ^{1}H, ^{13}C and ^{15}N signals of zinc-bound form of *Drosophila* GCM protein. Very recently, the crystal structure of GCM-box of mouse GCMa bound to DNA, which has 55% sequence identity to *Drosophila* GCM, has been deposited to PDB (PDB code: 1ODH). Combined with that, our data can provide a basis for structural study of the DNA recognition by the GCM-box, as well as the identification of the cysteine residues coordinating zinc atoms.

Methods and experiments

A recombinant protein consisting of the residues Ala25 to Ser181 of *Drosophila* GCM was expressed as a thioredoxin fusion protein in *E. coli* strain BL21/DE3 (pLysS) (Novagene, Co.) using an in-house derivative of pGEMEX-1 vector system (Promega, Madison, WI, USA), together with 'His-patch' thioredoxin gene (Invitorgen) and PreScission protease recognition sequence. The fusion protein was then purified with affinity chromatography using a Ni^{2+} charged HiTrap chelating column (Amersham Pharmacia Biotech, Uppsala, Sweden) and digested by PreScission protease (Amersham Pharmacia Biotech). To separate the N-terminal domain of GCM from

*To whom correspondence should be addressed. E-mail: hiroakih@tsurumi.yokohama-cu.ac.jp

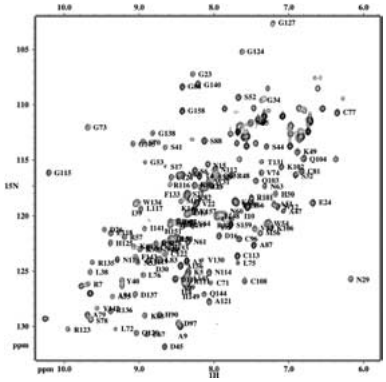

Figure 1. A portion of the ^{15}N-HSQC spectrum of the N-terminal domain of *Drosophila* GCM protein illustrating a number of the assigned backbone ^{15}N resonances.

the digest, cation exchanging column chromatography was used with a HiTrap SP column (Amersham Pharmacia Biotech) under 0–500 mM KCl salt gradient. Since two additional residues from the expression vector present, Ala25 in *Drosophila* GCM corresponds to Ala3 in the numbering system used here.

^{15}N uniformly labeled protein sample was prepared by growing cells on M9 minimal media containing $^{15}NH_4Cl$ as a sole nitrogen source. ^{15}N-uniform and ^{14}N-selective inversely labeled samples were also prepared using M9 minimal media containing $^{15}NH_4Cl$ supplemented with 100 mg l^{-1} of selected combinations of each amino acids (manuscript in preparation). Uniformly $^{13}C/^{15}N$-labeled sample was prepared by growing cells in BioExpress-Min medium (Cambridge Isotope Laboratory, Andover, MA, USA).

Extensive testing of solution conditions revealed that the N-terminal domain of *Drosophila* GCM protein was most stable in presence of 300 mM KCl and 5 to 10% glycerol. However, even in this condition, the samples were gradually precipitating at protein concentration higher than 0.6 mM during 3D NMR measurements. NMR samples contained up to 0.6 mM protein were in 99.9 % D_2O or 95 % H_2O – 5 % D_2O buffer containing 20 mM potassium phosphate, pH 6.5, 300 mM KCl and 10 % glycerol. All NMR spectra were recorded at 293 K on Bruker DMX spectrometers working at 600 or 750 MHz proton frequency equipped with Bruker TXI three axis gradient ^{1}H {^{13}C, ^{15}N} probes.

$^{1}H,^{15}N$-HSQC (Figure 1), HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, HNCO and HN(CA)CO spectra were recorded to obtain an assignment of backbone signals (Yamazaki et al., 1994). Spectra were processed using NMRPipe (Delaglio et al. 1995) and analyzed using XEASY (Bartels et al., 1995). Semi-automatic assignment method and further confirmation of assigned signals were then carefully carried out by XEASY powered by in-house program BASIL (Hiroaki, manuscript in preparation) using the combinatorial residue-specific ^{14}N inversely labeled samples. Finally, further confirmation of assignment was obtained by sequential NOEs in the $^{1}H,^{15}N$-NOESY-HSQC spectrum.

Extent of assignments and data deposition

The use of the residue-specific ^{14}N inversely labeled protein samples helped to clarify and to confirm a number of tentative assignments. In total, 96 % of the ^{1}HN and ^{15}N resonances of backbone amides (144 out of 150 possible) were assigned, and 98 % and 94% of the $^{13}C\alpha$ and ^{13}CO resonances were assigned. Complete backbone assignments of [N58, T59, H62, G124, Y128, G147] were not obtained. However, all cysteine residues were assigned, which involve all possibly zinc-binding cysteines. The backbone shifts have been deposited in the BioMagResBank under accession number 5626

Acknowledgements

The authors wish to thank Ms M. Yamaguchi and Prof H. Toh for their help in programming of in-house software. This work was partly supported by the Grant-in-Aid form Japan Society for the Promotion of Science (No. 12780498).

References

Akiyama, Y., Hosoya, T., Poole, A.M. and Hotta, Y. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 14912–14916.

Bartels, C., Xia, T., Billeter, M., Guentert, P. and Wüthrich K. (1995) *J. Biomol. NMR*, 6, 1–10.

Cohen, S.X., Moulin, M., Schilling, O., Meyer-Klaucke, W., Schreiber, J., Wegner, M. and Muller, C.W. (2002) *FEBS Lett.*, **528**, 95–100.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

Hosoya, T., Takizawa, K., Nitta, K. and Hotta, Y. (1995) *Cell*, **82**, 1025–1036.

Jones, B.W., Fetter, R.D., Tear, G. and Goodman, C.S. (1995) *Cell*, **82**, 1013–1023.

Schreiber, J., Enderich, J. and Wegner, M. (1998) *Nucl. Acids Res.*, **26**, 2337–2343.

Tuerk, E.E., Schreiber, J. and Wegner, M. (2000) *J. Biol. Chem.*, **275**, 4774–4782.

Wegner, M. and Riethmacher, D. (2001) *Trends Genet.* **17**, 286–290.

Yamazaki, T., Lee, W., Arrowsmith, C.H., Muhandiram, D.R. and Kay, L.E. (1994) *J. Am. Chem. Soc.*, **116**, 11655–11666.