Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Crystallization and preliminary X-ray diffraction analysis of mouse galectin-4 N-terminal carbohydrate recognition domain in complex with lactose

Veronika Krejčiříková,[a] Milan Fábry,[a] Vladimíra Marková,[a] Petr Malý,[a] Pavlína Řezáčová[a,b] and Jiří Brynda[a,b]*

[a]Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 37 Praha 6, Czech Republic, and [b]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 10 Praha 6, Czech Republic

Correspondence e-mail: brynda@img.cas.cz

Received 18 December 2007
Accepted 9 June 2008

Galectin-4 is thought to play a role in the process of tumour conversion of cells of the alimentary tract and the breast tissue; however, its exact function remains unknown. With the aim of elucidating the structural basis of mouse galectin-4 (mGal-4) binding specificity, we have undertaken X-ray analysis of the N-terminal domain, CRD1, of mGal-4 in complex with lactose (the basic building block of known galectin-4 carbohydrate ligands). Crystals of CRD1 in complex with lactose were obtained using vapour-diffusion techniques. The crystals belong to tetragonal space group $P42_12$ with unit-cell parameters a = 91.1, b = 91.16, c = 57.10 Å and preliminary X-ray diffraction data were collected to 3.2 Å resolution. An optimized crystallization procedure and cryocooling protocol allowed us to extend resolution to 2.1 Å. Structure refinement is currently under way; the initial electron-density maps clearly show non-protein electron density in the vicinity of the carbohydrate binding site, indicating the presence of one lactose molecule. The structure will help to improve understanding of the binding specificity and function of the potential colon cancer marker galectin-4.

1. Introduction

Galectins, a family of evolutionarily conserved β-galactoside binding proteins (Barondes *et al.*, 1994), can bind various galactose-containing oligosaccharides and thus play a role in important processes such as mitosis, apoptosis and cell-cycle progression (Cooper, 2002). Although galectins do not have any protein secretion signal sequence, they are localized not only in intracellular compartments, but also on the cell surface (Ackerman *et al.*, 2002). The mechanism of their extracellular targeting is not known (Hughes, 2001). The conserved galectin family is defined by the presence of a homologous carbohydrate recognition domain (CRD) with µM–nM affinity for various β-galactosides (Barondes *et al.*, 1994; Kilpatrick, 2002; Stowell *et al.*, 2008). These features are common to all 14 members of the galectin family. Based on the domain composition, the galectin family can be divided into three subfamilies: (1) chimera type (galectin-3), (2) monomer type (galectins-1, 2, 5, 7, 10, 11, 13, 14) (Lovell *et al.*, 2003), and (3) tandem-repeat subfamily (galectins-4, 6, 8, 9, 12) (Wang *et al.*, 2004).

Galectin-4 is composed of two distinct but homologous domains, CRD1 and CRD2. These domains share 40% identity and are connected by a proline- and glycine-rich peptide linker (Huflejt *et al.*, 1997), which is highly sensitive to tissue proteases. In fact, galectin-4 was first discovered as a 17 kDa protein in rat intestinal extract (Oda *et al.*, 1993); it was only after gene cloning that Oda *et al.* found that this protein was a proteolytic fragment of a larger 36 kDa protein.

Several studies have confirmed the expression of galectin-4 on the surface of the epithelium lining of the whole alimentary tract (Danielsen & van Deurs, 1997). The lipid raft microdomains of intestinal epithelial cells contain galectin-4 as an organizer and stabilizer of the brush border membrane, preventing loss of digestive enzymes and protecting the glycolipid microdomains (Danielsen & Hansen, 2006; Wrackmeyer *et al.*, 2006). Galectin-4 was also detected in the lung, breast, liver and placenta by the highly sensitive techniqe of RT-PCR (reverse-transcription polymerase-chain reaction) (Nagy *et al.*, 2003). Ideo *et al.* (Ideo *et al.*, 2002, 2005) recently demonstrated

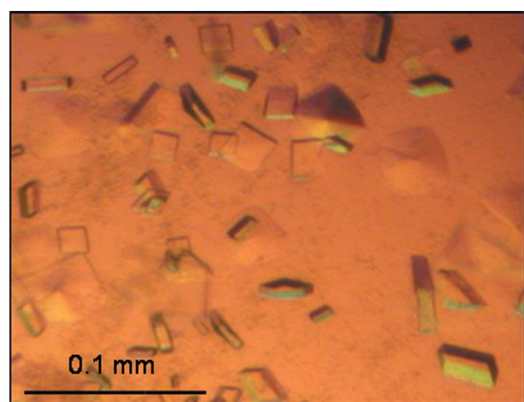
© 2008 International Union of Crystallography
All rights reserved

that galectin-4 binds to O-linked sulfoglycans and to sulfated glycosphingolipids, and to carcinoembryonic antigen present on the cell surface of human colon adenomacarcinoma cells. Galectin-4 may thus be a valuable diagnostic marker and target for the development of inhibitory carbohydrate-based drugs (Huflejt & Leffler, 2004).

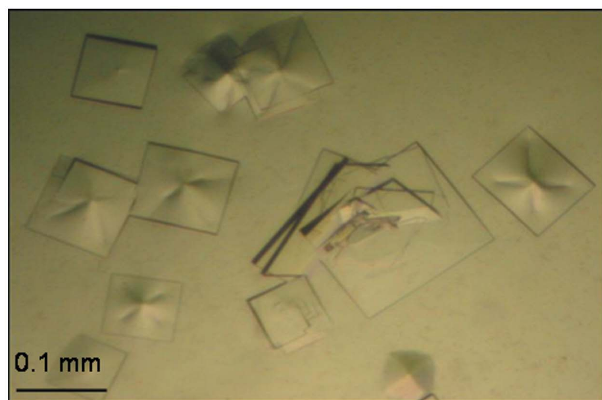
The presence of two CRD non-identical core domains separated by a link region suggests the possible role of tandem-repeat galectin members in a crosslinking function (Brewer, 2002). It is believed that mGal-4 could crosslink two distinct types of ligands, rather than identical molecules. Therefore, binding specificity for each domain was recently studied in rat and human galectin-4 (Sorme *et al.*, 2003; Holikova *et al.*, 2002; Cao *et al.*, 2002). The function of galectin-4 still remains unclear, however.

Several crystal structures of human galectins have already been solved and deposited in the PDB, *e.g.* Leonidas *et al.* (1998) who reported the high-resolution crystal structure of human galectin-7 (hGal-7), in free form and in the presence of galactose, galactosamine, lactose and *N*-acetyl-lactosamine.

In this work we report the crystallization of the N-terminal carbohydrate recognition domain of mouse galectin-4 (mGal-4) in complex with lactose. An optimized crystallization procedure yielded crystals suitable for X-ray analysis, and high-quality diffraction data will be used to determine the structure of the complex and elucidate the structural basis of the carbohydrate binding specificity of mGal-4.



(*a*)



(*b*)

Figure 1
Crystals of galectin-4 CRD1 domain in complex with lactose. (*a*) Initial crystals obtained using the vapour-diffusion hanging-drop technique using reservoir solution 0.1 *M* sodium cacodylate buffer pH 5.0, 15%(*w*/*v*) PEG 4000, 0.2 *M* $(NH_4)_2SO_4$, 5 m*M* β-mercaptoethanol, 80 m*M* lactose. (*b*) Optimized crystals grown using the same reservoir solution listed above, supplemented with 18%(*v*/*v*) glycerol combined with introducing crystal nuclei by microseeding.

2. Materials and methods

2.1. Molecular cloning, protein expression and purification

The plasmid used for protein expression contained sequence coding for amino acids 2–142 of mouse galectin-4 N-terminal carbohydrate recognition domain (the GenBank accession number of the complete mGal-4 sequence is *AY044870*). This sequence was preceded by a segment coding for the MRGSHHHHHHTDP affinity tag present in the pQE-31 expression vector (QIAGENE). A detailed description of the molecular cloning and construction of the bacterial expression vector for the N-terminal CRD1 domain of mGal-4 has been published previously (Markova *et al.*, 2006).

For expression, the *Escherichia coli* M15 cells containing the expression plasmid were cultivated in LB media at 310 K until the culture reached A_{550nm} = 1. The temperature was then lowered to 303 K, protein expression was induced with 1 m*M* IPTG (isopropyl-β-D-1-thiogalactosylpyranoside) (at A_{550nm} = 1.2) and the culture grown for 4 h.

Bacterial cells were disrupted by sonication in PBS (phosphate buffered saline) buffer (10 m*M* sodium phosphate buffer pH 7.4, 140 m*M* NaCl, 5 m*M* EDTA, 5 m*M* β-mercaptoethanol) supplemented with 50 µ*M* PMSF (phenylmethanesulfonyl fluoride). Cell supernatant was applied on a lactosyl–agarose (Sigma Aldrich) column at 277 K. The column was washed with 40 column volumes of PBS buffer and the bound CRD1 domain was then eluted by PBS buffer supplemented with 200 m*M* lactose. The flow rate was maintained at 0.1 ml min^{-1} during all steps. Final purification was achieved by gel filtration on Superdex 200 HR (10/30) (Amersham Biosciences) in PBS buffer. Purified protein migrated on SDS–PAGE as one clear band whose position corresponded well to the calculated molecular weight of 18.75 kDa. Purified protein in PBS buffer was stored at 203 K.

2.2. Crystallization

For crystallization, mGal-4 CRD1 domain in PBS buffer was concentrated to 9.6 mg ml^{-1} using ultrafiltration. Crystallization experiments were performed at 291 K by the hanging-drop vapour-diffusion method using classical 24-well Linbro boxes with 1 ml crystallization buffer in reservoir. Crystallization drops consisted of 1 µl of protein solution, 0.2 µl of 0.8 *M* lactose solution and 1 µl of reservoir solution. The final molar excess of lactose over protein in the drop was 300-fold.

2.3. Data collection

The first data set was collected on in-house X-ray equipment (MAR345 Image Plate Detector, Nonius FR 591 generator) using Cu *K*R1 radiation. Crystals were cryocooled in nitrogen gas stream (Oxford Cryosystems) at 110 K. A total of 93 diffraction images were collected, each covering 1° of crystal rotation. All data were integrated using *MOSFLM* (Leslie, 1999) and scaled using *SCALA* (Evans, 1993), both from the *CCP*4 package (Collaborative Computational Project, Number 4, 1994). A high-resolution data set was collected at 100 K at beamline 19-ID of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory, Argonne, IL, USA. A total of 111 diffraction images were collected, each covering 1° of crystal rotation. Diffraction data were processed using the *HKL-3000* suite of programs (Minor *et al.*, 2006).

Table 1
Crystal parameters and data-collection statistics.

Statistics for the highest-resolution shell are in parentheses.

X-ray source	Home source	19-ID, APS
Wavelength (Å)	1.54	0.9795
Temperature (K)	110.0	100.0
Space group	$P42_12$	$P42_12$
Unit-cell parameters: a, b, c (Å)	91.16, 91.16, 57.10	90.84, 90.84, 57.13
Resolution limit (Å)	35.62–3.20 (3.28–3.20)	35.62–2.10 (2.18–2.10)
R_{sym}†	0.241 (0.695)	0. 071 (0.472)
No. of unique reflections	4260 (302)	14,348 (1374)
Completeness	99.5 (100.0)	99.2 (97.5)
Multiplicity	6.1 (6.2)	7.6 (5.4)
$\langle I/\sigma(I)\rangle$	9.2 (2.8)	9.2 (2.4)
Wilson B (Å^2)	49.7	20.5

† $R_{sym} = \sum_{\mathbf{h}} \sum_l |I_{\mathbf{h}l} - \langle I_{\mathbf{h}}\rangle| / \sum_{\mathbf{h}} \sum_l \langle I_{\mathbf{h}}\rangle$ where I_l is the lth observation of reflection **h** and $\langle I_{\mathbf{h}}\rangle$ is the weighted average intensity for all observations l of reflection **h**.

3. Results and discussion

The first crystallization trials were performed using Crystal Screens 1 and 2 (Hampton Research). In the hanging-drop experiments, 1 µl of protein solution was mixed with 0.2 µl of 0.8 *M* lactose and 1 µl of reservoir solution, supplemented with 5 m*M* β-mercaptoethanol. Initial microcrystals of size approximately 30 × 30 × 5 µm were obtained in Crystal Screen 1 condition #15 [0.1 *M* sodium cacodylate pH 6.5, 15%(*w*/*v*) PEG 8000, 0.2 *M* $(NH_4)_2SO_4$] supplemented with 5 m*M* β-mercaptoethanol. By subsequent optimization, the quality of crystals was improved by replacing PEG 8000 with PEG 4000 and changing the pH to 5.0. The optimized crystallization procedure yielded crystals with maximal dimensions of 0.05 × 0.05 × 0.06 mm (Fig. 1*a*).

A complete set of diffraction data at 110 K from crystals soaked in 20% glycerol was collected up to 3.2 Å resolution on the in-house diffractometer (Nonius FR 591 generator using Cu $K\alpha_1$ radiation with Mar345 Image Plate Detector). Crystal parameters and data-collection statistics are summarized in Table 1. Our attempts to solve the phase problem by molecular replacement were successful when the protein chain from the structure of human galectin-7 (PDB code 2gal) (Leonidas *et al.*, 1998) was used as a search model. The tetragonal crystal form (space group $P42_12$) contained one molecule in the asymmetric unit with a solvent content of 61.2% (V_M = 3.17 Å^3 Da^{-1}).

The quality of the diffraction data obtained at the home source was limited but obtaining higher-quality diffraction data was, to some extent, hindered by the extreme sensitivity of the crystals to various cryoprotectants. To overcome this problem, attempts were made to introduce suitable cryoprotectant directly into the crystallization conditions. The addition of 18%(*v*/*v*) glycerol into optimized crystallization condition [0.1 *M* sodium cacodylate buffer pH 5.0, 15%(*w*/*v*) PEG 4000, 0.2 *M* $(NH_4)_2SO_4$, 5 m*M* β-mercaptoethanol, 0.2 m*M* lactose] did not prevent crystal growth; however, it slowed down crystal nucleation and growth quite significantly (by up to 2 months). To accelerate the nucleation event, microseeding (strike seeding) was further implemented and an optimized crystallization procedure yielded crystals with dimensions of 0.1 × 0.1 × 0.06 mm within 2 weeks (Fig. 1*b*).

Crystals grown in the presence of cryoprotectant could be cryo-cooled directly in their mother liquor, thus preventing crystal damage and stress from crystal soaking. These crystals diffracted to minimum Bragg spacing (d_{min}) 2.1 Å using the high brilliance synchrotron source (19-ID at APS Argonne, IL). Crystal parameters and data-collection statistics are summarized in Table 1. These crystals were isomorphous with the crystals grown previously in the absence of glycerol.

Structure refinement is currently under way. However, the initial electron-density maps clearly show non-protein electron density in the vicinity of the carbohydrate binding site, indicating the presence of one lactose molecule. The structure will help to improve understanding of the binding specificity and function of the colon cancer marker galectin-4.

This work was supported by the Academy of Sciences of the Czech Republic, grant Nos. 304/03/0090, AV0Z50520514 and AV0Z40550506. The authors wish to thank members of the Structural Biology Center at Argonne National Laboratory for their help with conducting data collection at the 19-ID beamline.

References

Ackerman, S. J., Liu, L., Kwatia, M. A., Savage, M. P., Leonidas, D. D., Swaminathan, G. J. & Acharya, K. R. (2002). *J. Biol. Chem.* **277**, 14859–14868.
Barondes, S. H., Castronovo, V., Cooper, D. N., Cummings, R. D., Drickamer, K., Feizi, T., Gitt, M. A., Hirabayashi, J., Hughes, C., Kasai, K. *et al.* (1994). *Cell*, **76**, 597–598.
Brewer, F. C. (2002). *Biochim. Biophys. Acta*, **1572**, 255–262.
Cao, Z., Said, N., Amin, S., Wu, H. K., Bruce, A., Garate, M., Hsu, D. K., Kuwabara, I., Liu, F. T. & Panjwani, N. (2002). *J. Biol. Chem.* **277**, 42299–42305.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cooper, D. N. (2002). *Biochim. Biophys. Acta*, **1572**, 209–231.
Danielsen, E. M. & Hansen, G. H. (2006). *Mol. Membr. Biol.* **23**, 71–79.
Danielsen, E. M. & van Deurs, B. (1997). *Mol. Biol. Cell*, **8**, 2241–2251.
Evans, P. R. (1993). *Data reduction. Proceedings of the CCP4 Study Weekend*. CLRC Daresbury Laboratory, Warrington, England.
Holikova, Z., Hrdlickova-Cela, E., Plzak, J., Smetana, K. Jr, Betka, J., Dvorankova, B., Esner, M., Wasano, K., Andre, S., Kaltner, H., Motlik, J., Hercogova, J., Kodet, R. & Gabius, H. J. (2002). *APMIS*, **110**, 845–856.
Huflejt, M. E., Jordan, E. T., Gitt, M. A., Barondes, S. H. & Leffler, H. (1997). *J. Biol. Chem.* **272**, 14294–14303.
Huflejt, M. E. & Leffler, H. (2004). *Glycoconj. J.* **20**, 247–255.
Hughes, R. C. (2001). *Biochimie*, **83**, 667–676.
Ideo, H., Seko, A., Ohkura, T., Matta, K. L. & Yamashita, K. (2002). *Glycobiology*, **12**, 199–208.
Ideo, H., Seko, A. & Yamashita, K. (2005). *J. Biol. Chem.* **280**, 4730–4737.
Kilpatrick, D. C. (2002). *Biochim. Biophys. Acta*, **1572**, 187–197.
Leonidas, D. D., Vatzaki, E. H., Vorum, H., Celis, J. E., Madsen, P. & Acharya, K. R. (1998). *Biochemistry*, **37**, 13930–13940.
Leslie, A. G. W. (1999). *Acta Cryst.* D**55**, 1696–1702.
Lovell, S. C., Davis, I. W., Arendall, W. B. III, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.
Markova, V., Smetana, K. Jr, Jenikova, G., Lachova, J., Krejcirikova, V., Poplstein, M., Fabry, M., Brynda, J., Alvarez, R. A., Cummings, R. D. & Maly, P. (2006). *Int. J. Mol. Med.* **18**, 65–76.
Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. (2006). *Acta Cryst.* D**62**, 859–866.
Nagy, N., Legendre, H., Engels, O., Andre, S., Kaltner, H., Wasano, K., Zick, Y., Pector, J. C., Decaestecker, C., Gabius, H. J., Salmon, I. & Kiss, R. (2003). *Cancer*, **97**, 1849–1858.
Oda, Y., Herrmann, J., Gitt, M. A., Turck, C. W., Burlingame, A. L., Barondes, S. H. & Leffler, H. (1993). *J. Biol. Chem.* **268**, 5929–5939.
Sorme, P., Kahl-Knutson, B., Wellmar, U., Nilsson, U. J. & Leffler, H. (2003). *Methods Enzymol.* **362**, 504–512.
Stowell, S. R., Qian, Y., Karmakar, S., Koyama, N. S., Dias-Baruffi, M., Leffler, H., McEver, R. P. & Cummings, R. D. (2008). *J. Immunol.* **180**, 3091–3102.
Wang, J. L., Gray, R. M., Haudek, K. C. & Patterson, R. J. (2004). *Biochim. Biophys. Acta*, **1673**, 75–93.
Wrackmeyer, U., Hansen, G. H., Seya, T. & Danielsen, E. M. (2006). *Biochemistry*, **45**, 9188–9197.