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

Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of the N-terminal carbohydrate-recognition domain of human galectin-4

Ana Lucia L. R. Zimbardi,[a] Matheus P. Pinheiro,[a] Marcelo Dias-Baruffi[b] and M. Cristina Nonato[a]*

[a]Laboratório de Cristalografia de Proteínas, Departamento de Física e Química, Faculdade de Ciências Farmacêuticas de Ribeirão Preto - USP, Ribeirão Preto - SP, 14040-903, Brazil, and [b]Laboratório de Glicoimunobiologia, Departamento de Análises Clínicas Toxicológicas e Bromatológicas, Faculdade de Ciências Farmacêuticas de Ribeirão Preto - USP, Ribeirão Preto - SP, 14040-903, Brazil

Correspondence e-mail: cristy@fcfrp.usp.br

Received 18 September 2009
Accepted 23 March 2010

© 2010 International Union of Crystallography
All rights reserved

Galectin-4 is a tandem-repeat-type galectin that is expressed in the epithelium of the alimentary tract from the tongue to the large intestine. Additionally, strong expression of galectin-4 can also be induced in cancers in other tissues, including the breast and liver. In order to explore its potential as a target for anticancer drug design, elucidation of the structural basis of the carbohydrate-binding specificities of galectin-4 has been focused on. As an initial step, the N-terminal carbohydrate-recognition domain of human galectin-4 (hGal4-CRD-1) has been successfully crystallized using the vapour-diffusion technique, a complete data set has been collected to 2.2 Å resolution and the structure has been solved by the molecular-replacement technique. The crystals belonged to space group $P6_122$, with unit-cell parameters $a = b = 71.25$, $c = 108.66$ Å. The asymmetric unit contained one molecule of hGal4-CRD-1, with a V_M value of 2.34 Å^3 Da^{-1} and a solvent content of 47.51%.

1. Introduction

Galectins are a family of animal lectins that recognize β-galactosides. They are ubiquitously expressed and distributed in mammalian tissues and are involved in fundamental cellular processes such as growth regulation, adhesion/migration and apoptosis (Cooper, 2002). More recently, galectins have also been shown to bind glycans on the surface of potentially pathogenic microorganisms and to function as recognition and effector factors in innate immunity (Vasta, 2009).

All galectins share a conserved carbohydrate-recognition domain (CRD) consisting of about 130 amino-acid residues, many of which are highly conserved (Cooper & Barondes, 1999). The architecture of the CRDs in galectins of known three-dimensional structure is very similar and is composed of an 11-strand or 12-strand antiparallel β-sandwich folded in a jelly-roll topology typical of legume lectins (Bianchet *et al.*, 2000; López-Lucendo *et al.*, 2004; Liao *et al.*, 1994; Lobsanov *et al.*, 1993; Walser *et al.*, 2004; Seetharaman *et al.*, 1998; Leonidas *et al.*, 1998; Nagae *et al.*, 2008; Ackerman *et al.*, 2002; Zhou *et al.*, 2008).

Based on their domain organization, galectins can be classified into three subfamilies: prototype, chimera type and tandem-repeat type. Prototype galectins contain one CRD per subunit and are usually homodimers (López-Lucendo *et al.*, 2004). The chimera type contain a C-terminal CRD linked to a long peptide rich in proline and glycine residues (Seetharaman *et al.*, 1998). The tandem-repeat-type galectins are composed of two carbohydrate-recognition domains that are located in the N- and C-terminal portions of the protein (CRD-1 and CRD-2, respectively) and connected by a linker peptide (Jiang *et al.*, 1999).

Human galectin-4 (hGal4), a 36 kDa tandem-repeat-type galectin, is expressed in the epithelium of the alimentary tract from the tongue to the large intestine. In addition, differential expression of hGal4 can be induced in cancers in other tissues, including the breast and liver (Huflejt & Leffler, 2004), supporting the idea that hGal4 can be used as a prognostic indicator of tumour progression in pathological analysis and as a target for anticancer drug design.

Although both CRD-1 and CRD-2 of hGal4 present the characteristic amino-acid sequence for β-galactoside recognition and binding (Cooper & Barondes, 1999), they share only 40% sequence identity. Both domains bind lactose with similar affinity to other

galectins, but their preferences for other disaccharides and larger saccharides differ distinctly (Huflejt & Leffler, 2004). Owing to the presence of the linking region connecting the two CRDs, it is believed that hGal4 functions as a crosslinker of two distinct types of ligands (Brewer, 2002).

We are interested in elucidating the structural basis of the carbohydrate-binding specificities observed for each CRD in galectin-4. Currently, the structural information available for galectin-4 is limited to a crystallographic structure of the N-terminal domain of mouse galectin-4 (PDB code 2dyc; M. Kato-Murayama, K. Murayama, T. Terada, M. Shirouzu & S. Yokoyama, unpublished work) and a solution structure of the C-terminal domain of hGal4 (PDB code 1x50; T. Tomizawa, T. Kigawa, K. Saito, S. Koshiba, M. Inoue & S. Yokoyama, unpublished work). Crystallization of the N-terminal domain of mouse galectin-4 in complex with lactose has also recently been reported (Krejčiříková *et al.*, 2008).

In the present work, we report the overexpression, crystallization and preliminary X-ray diffraction analysis of the N-terminal carbohydrate-recognition domain of human galectin-4 (hGal4-CRD-1) in its apo form. The work presented here constitutes the first step towards the structure determination of hGal4-CRD-1 as a significant contribution towards a better understanding of the role of galectin-4 in normal and pathological conditions.

2. Methods

2.1. Cloning

The gene encoding full-length hGal4, which was kindly provided by Professor Richard Cummings of Emory University School of Medicine, was used as a template for the amplification process. The forward primer 5′-GACGACCATATGGCCTATGTCCCCGCAC-3′ and the reverse primer 5′-GACGACGCGGCCGCATTAGCCTCC-GAGAAGTTGATTGATTGAAG-3′ were designed to amplify the gene sequence coding for amino acids 1–150 of hGal4 (NCBI Reference Sequence NP_006140.1) and to introduce *Nde*I and *Not*I restriction sites. The polymerase chain reaction (PCR) was carried out in a 50 µl volume containing 0.4 µ*M* of each primer, 120 ng DNA template, 0.2 m*M* dNTP mixture, 1 U Platinum TAQ polymerase High Fidelity (Invitrogen) and 2 m*M* $MgSO_4$. The reaction was performed by initial denaturation at 367 K for 120 s followed by 25 cycles of 367 K for 30 s (denaturation), 328 K for 30 s (annealing) and 345 K for 70 s (elongation). Final extension was performed at 345 K for 10 min. The resulting PCR product was visualized on a 0.8% agarose gel stained with ethidium bromide, gel extracted and purified using a QIAquick Gel Extraction Kit (Qiagen). The purified PCR product was ligated into pTZ57R/T vector (Fermentas) using T4 DNA ligase (Invitrogen) and transformed into DH5α competent cells for plasmid propagation. Plasmid DNA was extracted from the transformants using a QIAprep Spin Miniprep Kit (Qiagen) and the sequence of the cloned gene was confirmed by DNA sequencing. The hGal4-CRD-1 gene fragment was digested from the pTZ57R/T vector construct with *Nde*I and *Not*I (Invitrogen) to provide cohesive ends, gel purified and ligated to digest the pET-28a (Novagen) expression vector using T4 DNA ligase. The ligated mixture was transformed into DH5α competent cells for plasmid propagation and transformed into BL21 (DE3) cells for expression and purification.

2.2. Expression and purification

hGal4-CRD-1 was expressed as a His_6-tag fusion protein. A single colony of BL21 (DE3) cells harbouring the expression construct was inoculated into 5 ml Luria broth (LB) medium containing 50 µg ml^{-1} kanamycin and allowed to grow overnight at 310 K with shaking. The overnight culture was then diluted 1:100 with LB containing 50 µg ml^{-1} kanamycin and grown at 310 K until the OD_{600} reached 0.4–0.5. The cells were induced with 1 m*M* isopropyl β-D-1-thio-galactopyranoside (IPTG) and growth continued for 24 h at 303 K. The induced culture was harvested by centrifugation at 5000*g* for 20 min. An *E. coli* pellet was suspended in 6 ml lysis buffer (50 m*M* sodium phosphate pH 8, 300 m*M* NaCl and 1 m*M* phenylmethylsulfonyl fluoride) and the cell suspension was lysed by sonication in 8 × 15 s bursts (with 15 s intervals between each burst) using a Misonix XL 2000 sonicator fitted with a microtip probe and set at power setting 5. The crude extract was clarified by centrifugation at 10 000*g* for 30 min. The hGal-4 N-terminal fragment found in the soluble fraction was loaded onto a 2 ml column of Ni–NTA resin equilibrated with lysis buffer (50 m*M* sodium phosphate pH 8, 300 m*M* NaCl). The column was washed with a step gradient of 0–50 m*M* imidazole in lysis buffer (50 m*M* sodium phosphate pH 8, 300 m*M* NaCl plus imidazole). The washes consisted of 40 ml lysis buffer followed by 30 ml 25 m*M* imidazole in lysis buffer and 20 ml 50 m*M* imidazole in lysis buffer. The recombinant hGal4-CRD-1 was eluted with approximately 20 ml 100 m*M* imidazole solution (50 m*M* sodium phosphate pH 8, 300 m*M* NaCl and 100 m*M* imidazole) and the protein purity was checked by 12% SDS–PAGE (Fig. 1). The purified protein migrated on SDS–PAGE as one clear band, the position of which corresponded well to the calculated molecular weight of 17 kDa. The concentration of the purified protein was determined using UV absorbance spectroscopy at 280 nm, which estimates the protein concentration based on the molar extinction coefficients of tryptophan, tyrosine and cysteine residues (Pace & Schmidt, 1997). This method gave a calculated extinction coefficient for the recombinant protein of 24 410 M^{-1} cm^{-1}. Prior to crystal-

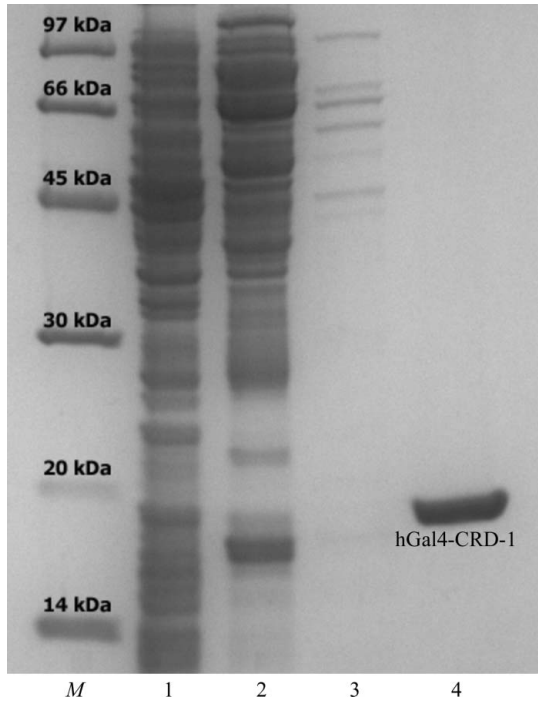

Figure 1
Analysis of the purification steps of hGal4-CRD-1 (12% SDS–PAGE stained with Coomassie Brilliant Blue). Lane 1, unbound fraction after passage over the affinity column; lane 2, sample of fraction washed with 25 m*M* imidazole; lane 3, sample of fraction washed with 50 m*M* imidazole; lanes 4; sample of the fraction eluted in the presence of 100 m*M* imidazole, representing the purified hGal4-CRD-1. Lane *M*, molecular-weight markers (kDa).

lization, recombinant hGal4-CRD1 was dialyzed against 50 m*M* HEPES pH 7.2, 150 m*M* NaCl and concentrated to 8.8 mg ml^{-1} (Millipore Ultrafree membrane; 10 kDa molecular-mass cutoff). The crystallization trials were carried out using His_6-tagged enzyme.

2.3. Crystallization and data collection

Initial crystallization conditions were screened by the sparse-matrix method (Jancarik & Kim, 1991) using commercially available screening kits and the sitting-drop vapour-diffusion technique. Equal volumes (2 µl) of protein and reservoir solution were mixed, equilibrated against 500 µl of reservoir solution and kept at 295 K. Small needles appeared within 4 d in two different conditions: 0.1 *M* sodium cacodylate trihydrate pH 6.5, 18%(*w*/*v*) PEG 8000, 0.2 *M* calcium acetate hydrate (Crystal Screen 1 formulation No. 46; Hampton Research) and 0.1 *M* Tris–HCl pH 8.5, 30%(*w*/*v*) PEG 4000, 0.2 *M* sodium acetate trihydrate (Crystal Screen formulation No. 22; Hampton Research). Several efforts were made to optimize the crystal quality by screening various crystallization variables (pH, polyethylene glycol size and concentration, temperature and additives). Crystals suitable for X-ray diffraction measurements were obtained in the presence of 0.1 *M* Tris–HCl pH 8.5, 16%(*w*/*v*) PEG 8000, 0.2 *M* calcium acetate hydrate. The crystals of hGal4-CRD-1 usually appeared within 4 d and reached maximum dimensions of 0.4 × 0.1 × 0.1 mm after 10 d (Fig. 2).

Single crystals of hGal4-CRD-1 were transferred to a cryoprotectant solution consisting of 0.1 *M* Tris–HCl pH 8.5, 16%(*w*/*v*) PEG 8000, 0.2 *M* calcium acetate hydrate, 20%(*v*/*v*) glycerol and flash-cooled directly in a nitrogen stream at 100 K. Data collection was performed on the D03B-MX1 beamline of the Brazilian Synchrotron Light Laboratory (LNLS) using radiation of wavelength 1.437 Å; data were recorded on a MAR CCD 165 detector. An exposure time of 60 s per image was used, with a crystal-to-detector distance of 90 mm. A total of 38 1.0° diffraction images were collected. The data were processed using *MOSFLM* (Leslie, 2006) and scaled using *SCALA* (Evans, 2006). The data-collection and processing statistics are summarized in Table 1.

2.4. Phase calculation and refinement

Initial phases were obtained by molecular replacement with the program *MOLREP* (Vagin & Teplyakov, 2000). Initial rotation and translation functions were calculated using the coordinates of a homologous N-terminal CRD from mouse galectin-4 (mGal4-CRD-1; PDB code 2dyc; M. Kato-Murayama, K. Murayama, T. Terada, M. Shirouzu & S. Yokoyama, unpublished work) as the search model. hGal4-CRD-1 and mGal4-CRD-1 share 81% sequence identity. The model was initially refined by rigid-body refinement, which was followed by several rounds of adjusting side-chain rotamers for residues using *Coot* (Emsley & Cowtan, 2004) interspersed with positional and individual *B*-factor refinement using *REFMAC*5 (Murshudov *et al.*, 1997).

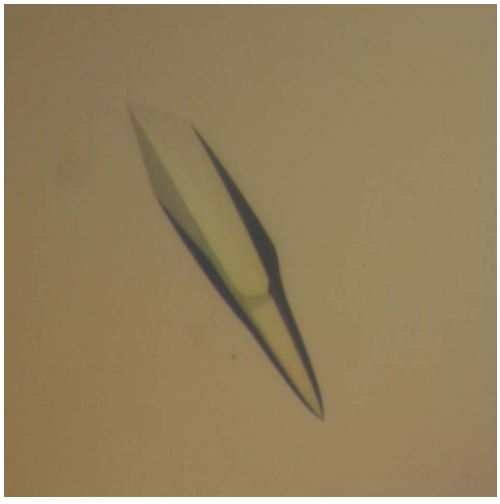

Figure 2
Crystals of hGal4-CRD-1.

Table 1
Data-collection and processing statistics.

Values in parentheses are for the outer resolution shell.

Temperature (K)	100
Wavelength (Å)	1.437
Space group	$P6_122$
Unit-cell parameters (Å)	a = 71.25, c = 108.66
Resolution range (Å)	40.79–2.2 (2.32–2.2)
Unique reflections	8809 (1250)
Redundancy	4.3 (4.4)
Completeness (%)	100 (100)
$\langle I/\sigma(I)\rangle$	10.1 (2.5)
$R_{\rm merge}$† (%)	11.8 (59.0)
Molecules per asymmetric unit	1
$V_{\rm M}$ (Å^3 Da^{-1})	2.34
Solvent content (%)	47.5

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean intensity of multiple observations of symmetry-related reflections.

3. Results and discussion

The crystals of apo hGal4-CRD-1 diffracted to 2.2 Å resolution. The unit-cell parameters (Table 1) and the systematic absences of the 00*l*, *l* = 6*n* reflections indicated that the crystals belonged to the hexagonal space group $P6_122$ or $P6_522$. A monomer of hGal4-CRD-1 has a calculated molecular weight of approximately 17 kDa; assuming the presence of one molecule in the asymmetric unit, the calculated Matthews coefficient was 2.34 Å^3 Da^{-1}. Based on a specific volume of 0.74 cm^3 g^{-1}, the calculated solvent content was approximately 47.5%, which lies within the range commonly observed for protein crystals (Matthews, 1968). The application of molecular-replacement procedures for both enantiomer possibilities indicated $P6_122$ as the correct space group, yielding a solution with a correlation coefficient of 46.1% and an *R* factor of 54.7%. Initial rigid-body refinement followed by 30 cycles of restrained refinement implemented with *REFMAC*5 (Murshudov *et al.*, 1997) resulted in R = 29.6% and $R_{\rm free}$ = 37.2%. Further structural refinement is in progress.

The crystallization and structure determination of hGal4-CRD-1 in complex with a variety of β-galactosides is currently in progress. The structural analysis of the ligand-binding regions will help to increase our understanding of the molecular mechanism for different binding specificities. The preparation of diffraction-quality crystals of apo hGal4-CRD-1 reported here represents an important first step towards this goal.

We would like to thank Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) and Conselho Nacional de Pesquisa (CNPq) for financial support.

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.

Bianchet, M. A., Ahmed, H., Vasta, G. R. & Amzel, L. M. (2000). *Proteins*, **40**, 378–388.
Brewer, F. C. (2002). *Biochim. Biophys. Acta*, **1572**, 255–262.
Cooper, D. N. W. (2002). *Biochem. Biophys. Acta*, **1572**, 209–231.
Cooper, D. N. W. & Barondes, S. H. (1999). *Glycobiology*, **9**, 979–984.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Huflejt, M. E. & Leffler, H. (2004). *Glycoconj. J.* **20**, 247–255.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Jiang, W., Guo, X. & Bhavanandan, P. (1999). *IUBMB Life*, **48**, 601–605.
Krejčiříková, V., Fábry, M., Marková, V., Malý, P., Řezáčová, P. & Brynda, J. (2008). *Acta Cryst.* F**64**, 665–667.
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. (2006). *Acta Cryst.* D**62**, 48–57.
Liao, D. I., Kapadia, G., Ahmed, H., Vasta, G. R. & Herzberg, O. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 1428–1432.
Lobsanov, Y. D., Gitt, M. A., Leffler, H., Barondes, S. H. & Rini, J. M. (1993). *J. Biol. Chem.* **268**, 27034–27038.
López-Lucendo, M. F., Solis, D., Andre, S., Hirabayashi, J., Kasai, K., Kaltner, H., Gabius, H. J. & Romero, A. (2004). *J. Mol. Biol.* **343**, 957–970.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Nagae, M., Nishi, N., Nakamura-Tsuruta, S., Hirabayashi, J., Wakatsuki, S. & Kato, R. (2008). *J. Mol. Biol.* **375**, 119–135.
Pace, C. N. & Schmidt, F. X. (1997). *Protein Structure: A Practical Approach*, edited by T. E. Creighton, pp. 253–259. New York: IRL Press.
Seetharaman, J., Kanigsberg, A., Slaaby, R., Leffler, H., Barondes, S. H. & Rini, J. M. (1998). *J. Biol. Chem.* **37**, 13047–13052.
Vagin, A. & Teplyakov, A. (2000). *Acta Cryst.* D**56**, 1622–1624.
Vasta, G. R. (2009). *Nature Rev. Microbiol.* **7**, 424–438.
Walser, P. J., Haebel, P. W., Kunzler, M., Sargent, D., Kues, U., Aebi, M. & Ban, N. (2004). *Structure*, **12**, 689–702.
Zhou, D., Ge, H., Sun, J., Gao, Y., Teng, M. & Niu, L. (2008). *Proteins*, **71**, 1582–1588.