Gene design, expression, crystallization and preliminary diffraction analysis of two isolectins from the fungus *Coprinus cinereus*: a model for studying functional diversification of galectins in the same organism and their evolutionary pathways

María F. López-Lucendo,[a]
Guillermo Giménez-Gallego,[a]
Douglas N. W. Cooper,[b]
Hans-Joachim Gabius[c] and
Antonio Romero[a]*

[a]Departamento de Estructura y Función de Proteínas, Centro de Investigaciones Biológicas–CSIC, C/Ramiro de Maeztu, 9, Madrid 28040, Spain, [b]Departments of Anatomy and Psychiatry, Langley Porter Psychiatric Institute, Center for Neurobiology and Psychiatry, University of California, 401 Parnassus Avenue, San Francisco, California 94143-0984, USA, and [c]Institut für Physiologische Chemie, Tierärztliche Fakultät, Ludwig-Maximilians-Universität München, Veterinärstrasse 13, D-80539 München, Germany

Correspondence e-mail: romero@cib.csic.es

It is the aim of comparative structural biology to define the evolutionarily important traits of protein function and the points of diversification. Consequently, structural analysis, especially of distant members in a family which in this case are lectins involved in cell adhesion and growth regulation in animals (*i.e.* galectins), is required. For this purpose, recent work has been focused on the first galectins known from outside the animal kingdom. These are the two isolectins from the basidiomycete *Coprinus cinereus* (inky cap mushroom), termed Cgl-1 and Cgl-2. Additionally, the close similarity (83% deduced amino-acid identity) but the pronounced difference in the expression patterns of these two fungal lectins during fruiting-body formation affords a suitable object for study of the relation of structural difference to the observed functional disparity in the same organism. Both galectins were crystallized after recombinant production. Crystals belong to either the orthorhombic space group $C222_1$ (Cgl-1) or the monoclinic space group $P2_1$ (Cgl-2). The latter crystals diffracted to 1.6 Å resolution using synchrotron radiation. To solve the phasing problem, a selenomethionine-containing variant of Cgl-1 was designed. Crystals isomorphous to those of the native counterpart were obtained. Their structural analysis will also be crucial to solving the structure of Cgl-2.

1. Introduction

The emerging importance of glycans of cellular glycoconjugates in serving as hardware for the storage and transfer of biological information, reflected by the coining of the term *sugar code*, has engendered an increasing interest in the study of the molecular details of protein–carbohydrate interactions (Gabius, 1998; Lis & Sharon, 1998; Solís *et al.*, 2001). In addition to insights into the fundamental mechanisms of this type of intermolecular recognition, the structural study of carbohydrate-binding proteins, especially lectins, is helpful in at least two further ways: to provide input for the design of custom-made pharmaceuticals targeting clinically relevant lectins and to define structural traits with the aim of determining evolutionary pathways and basic structure–function relationships (Rüdiger *et al.*, 2000; Loris, 2002). Our study intends to primarily address the second aim, focusing on members of a lectin family with widespread occurrence in the animal kingdom, *i.e.* galectins (Cooper, 2002). By binding to spatially accessible β-galactosides, mostly at the branch ends

of glycan chains, these lectins are known to serve a multitude of functions including mechanisms of cell adhesion or migration and growth regulation or immune-mediator release (Gabius, 1997, 2001; Hirabayashi, 1997; Kaltner & Stierstorfer, 1998; Rabinovich *et al.*, 2002). They also play a role in intracellular processes such as apoptosis signalling, micro-domain stabilization and pro-mRNA splicing (Liu *et al.*, 2002). Overall, these proteins are multifunctional by means of harbouring versatile modules to mediate these activities reliably.

The recent demonstration of the occurrence of galectins outside the animal kingdom is thus of considerable interest to guide further structural studies. New members of galectin genealogy were described from two edible fungi: *Agrocybe cylindracea* and the inky cap mushroom *Coprinus cinereus* (Cooper *et al.*, 1997; Yagi *et al.*, 1997; Hirabayashi *et al.*, 2002). In fact, their presence in fungi, which diverged from the main phylogenetic route even prior to plants and the Porifera as the oldest metazoa, set the emergence of galectins, which have already been documented in sponges and

nematodes, for example, back to at least a billion years ago (Hirabayashi *et al.*, 2002). What makes the case of the *C. cinereus* galectins (Cgl) particularly attractive is the fact that the two closely related prototype galectins [83% identity in deduced amino-acid sequence with only 26 differences, eight clearly being conservative, in the 16 408 Da (Cgl-1) and 16 671 Da (Cgl-2) proteins] are subject to conspicuously different regulation during fruiting-body formation (Boulianne *et al.*, 2000). Detection of Cgl-2 in the initial stage of this process in hyphal knots and of Cgl-1 exclusively later in fruiting mycelium and the tissue of fruiting bodies led to the conclusion to refer to them as developmental markers, their prominent extracellular localization being suggestive of involvement in hyphal–hyphal (cell–cell) interactions (Boulianne *et al.*, 2000). Two further observations deserve to be mentioned with regard to structural comparison: the absence of the Cys residues that have been implicated in activity maintenance in other family members and the partial dissociation of the Cgl-2 homodimer, which indicates a less strong interface association than in the mammalian prototype relatives (Cooper *et al.*, 1997; Kopitz *et al.*, 2003). An instructive case in this respect is human galectin-1. Its folding follows the jelly-roll-like motif and it is homodimeric with an intricate system of interprotomer interactions to achieve tight association (López-Lucendo *et al.*, unpublished results).

The reasons discussed above have prompted our work to crystallize the two closely related fungal galectins. In an evolutionary perspective, the roles of key residues in ligand binding and homodimer stability can be assessed. Also, clues to the explanation of the two tightly regulated courses of differential expression are expected to emerge from the detailed description of the topologies of the ligand-binding sites of the isolectins.

2. Materials and methods

2.1. Cloning, expression and purification

A synthetic gene based on the cDNA sequence of Cgl-1 as determined by Cooper *et al.* (1997) was designed with optimal codon usage for expression in *Escherichia coli*. The gene was constructed following the method of Di Donato *et al.* (1993), using eight partially overlapping oligonucleotides ranging from 60 to 70 bp in length. The full-length DNA sequence encoding Cgl-1 was divided into two separate fragments having a common *Bsu*36I site (Fig. 1). For the synthesis of the core template, the first two primers A (or A*) and B (or B*) were annealed and extended with DeepVent (New England Biolabs) DNA polymerase. The amplified DNAs of these PCR mixtures were applied as a template in a second elongation PCR run after addition of primers C (or C*) and D (or D*). Each fragment was cloned independently into the bacterial expression vector pUC19 and then cleaved with *Bsu*36I and *Bsa*I. The resulting *Bsu*36I- and *Bsa*I-flanked fragments were linked to form the final pUC19-Cgl-1 vector with unique cleavage sites for *Bsp*HI and *Hin*dIII in the product. After digestion, the 462 bp *Bsp*HI- and *Hin*dIII-flanked piece was cloned into pRAT4 (Peränen *et al.*, 1996) so that the protein could be expressed under the control of the T7 promoter. To exclude sequence deviations arising from this processing, DNA sequencing was performed. Protein was expressed in *E. coli* strain BL21 (DE3).

Cell cultures were grown in supplemented M9 minimal medium (Sambrook *et al.*, 1989) at 310 K until the A_{600nm} reached 0.7 and were then induced with a concentration of 0.5 mM isopropyl-β-D-thiogalactopyranoside. Cells were grown for a further period of 6 h to obtain a maximal yield of Cgl-1 and were then harvested by centrifugation. The pellets were resuspended in 10 mM phosphate buffer pH 7.0 containing 5 mM EDTA and lysed by sonication on ice. The supernatant was collected by centrifugation. The galectin was purified by affinity chromatography using an α-lactose-agarose column (Sigma) at 277 K. The resin was equilibrated with PBS and after washing to remove unbound material, the bound galectin was eluted with a linear gradient (0–0.2 M) of lactose in PBS. The purity of the sample was assessed by SDS–PAGE analysis.

Selenomethionine (SeMet) could not be incorporated into Cgl-1 because its sequence lacks methionine residues apart from the N-terminus. Identification of leucine residues in optimal positions to engineer suitable variants was performed by scrutinizing multiple amino-acid sequence alignment within the galectin family to preclude strictly conserved sites from alteration. Based on this comparison, the cDNA sequence was subjected to site-directed mutagenesis in positions 13, 42, 103 and 144. The gene that encodes SeMet-containing Cgl-1 was constructed using the same strategy as described for producing the cDNA of the native protein, as described above. SeMet-containing protein was produced in *E. coli* methionine auxotroph 834(DE3), grown in a defined medium containing MEM vitamins (Gibco, 1× final concentration), 0.2 mM seleno-DL-methionine (Sigma) and ampicillin (20 µg ml^{-1}). SeMet-containing Cgl-1 was purified as described above and the homogeneity of the protein was assessed by SDS–PAGE.

2.2. Crystallization and data collection

Crystallization was performed using the sitting-drop vapour-diffusion method at 295 K in a 24-well tissue-culture plate. Initial crystallization trials were conducted using commercially available screening kits from Hampton Research (Crystal Screen and Crystal Screen II). Each drop was prepared
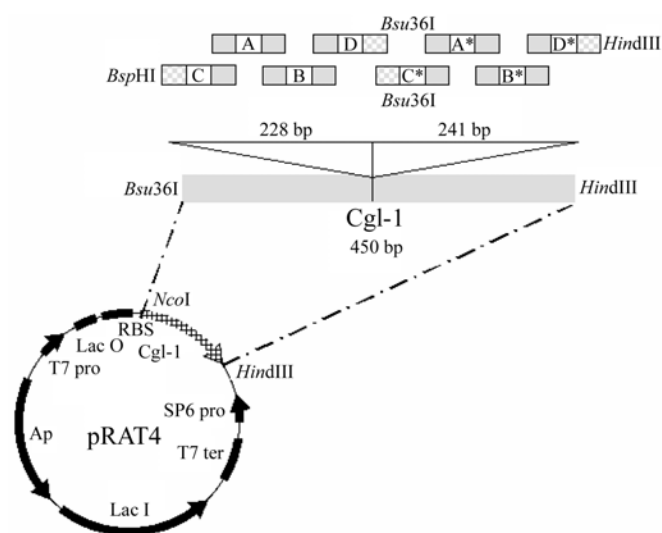

Figure 1
Construction of the 5′ Cgl-1 3′ plasmid into pRAT4. Partial overlapping using eight oligonucleotides was used to build the Cgl-1 gene according to the method of Di Donato *et al.* (1993).

by mixing 1 µl of protein (10 mg ml⁻¹) in 5 mM sodium phosphate, 150 mM NaCl, 0.1 M lactose pH 7.2 with the same volume of reservoir solution. Further improvement of the initially defined crystallization conditions was carried out by systematically exploring different precipitant concentrations, additives and buffer pHs. All diffraction data were collected at cryogenic temperatures. Before flash-freezing, crystals were briefly soaked in a solution derived from the mother liquor by adding 15% of PEG 400 as an additional cryoprotectant.

2.2.1. Native and SeMet Cgl-1. After 3 d, needle-shaped crystals were observed in one well from the Hampton Research Crystal Screen, which contained 30% PEG 4K at pH 8.5. Crystals of SeMet-containing Cgl-1 grew under similar conditions as observed for the native protein. However, optimization of the SeMet crystals was achieved with sodium fluoride, which promoted the most dramatic improvement in crystal size, morphology and diffraction quality. X-ray diffraction data from native Cgl-1 were obtained at ESRF beamline BM14 (Grenoble, France). A complete data set was collected from a single crystal to a resolution of 1.85 Å. Taking the subunit molecular weight to be

16.4 kDa suggests that the crystal contains a dimer in the asymmetric unit with a V_M of 2.5 Å³ Da⁻¹, which is within the range given by Matthews (1968). MAD diffraction data were acquired at EMBL beamline BW7A (DESY Hamburg, Germany) with a MAR CCD detector system. Three data sets were collected around the Se K edge to 2 Å resolution. Data were processed with *MOSFLM* v.6.0 (Leslie, 1997) and programs from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994).

2.2.2. Native Cgl-2. Briefly, Cgl-2 was purified by affinity chromatography on lactosyl-Sepharose as described previously (Cooper *et al.*, 1997). Small needles were obtained from several conditions containing PEGs, with drops assembled using 1 µl protein (10 mg ml⁻¹) in 5 mM sodium phosphate, 150 mM NaCl, 0.1 M lactose pH 7.2 and 1 µl reservoir solution. Conditions were optimized in a grid screen for buffer pH value (pH 4–6) and precipitants (PEG 1K, 2K, 3K and 4K). In the optimized crystallization condition, crystals were grown in 30% PEG 2K, 0.1 M MES pH 5.8. Crystallographic data were collected at the protein crystallography beamline BM14 at the ESRF (Grenoble, France). On the basis of

density calculations (V_M = 2.36 Å³ Da⁻¹; Matthews, 1968), we estimate that two dimers are present in the asymmetric unit. The diffraction data set was processed and scaled to 1.6 Å using *MOSFLM* and *SCALA* from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994).

3. Results and discussion

Crystals of native Cgl-1 and Cgl-2 formed readily using either PEG 4K or PEG 2K as the main precipitant, respectively (Fig. 2). They were either orthorhombic or monoclinic. For Cgl-1 crystals the space group was determined to be $C222_1$ on the basis of systematic absences, while Cgl-2 crystals belonged to space group $P2_1$. Data-processing statistics are presented in Table 1. To achieve structural characterization, we first attempted to take advantage of the structural relationship of the fungal galectin to animal galectins. However, attempts to elucidate the structures of Cgl-1 and Cgl-2 by molecular-replacement methods using the coordinates of several galectins, including human galectin-1 (PDB code 1gzw), as a starting model failed. Multiple sequence alignments yielded a score of 11–12% for both fungal galectins when compared with the sequences of human galectin-1 or the conger eel (*Conger myriaster*) galectin congerin I. For comparison, the human and the fish galectins shared 33% sequence identity. We next turned to the conventional multiple isomorphous replacement method. Numerous attempts to prepare heavy-atom derivatives were unsuccessful. Lack of isomorphism of the resulting crystals or a lack of definitive signals from isomorphous crystals that could locate heavy-atom positions forced us to move to a strategy involving protein engineering. Our choice was to incorporate selenomethionine to substitute leucine at distinct positions. This strategy follows the ranking given by the Dayhoff mutation probability (Jones *et al.*, 1992) based on comparisons of the amino-acid sequences of related proteins. It is also in accord with the observation that the lowest average isomorphous difference is attributed to a Leu to Met substitution, intimating that it constitutes the preferred choice for introducing SeMet in order to achieve MAD phasing (Gassner & Matthews, 1999).

To attain this aim in our case, the first time this strategy has been applied to a galectin, mutated Cgl-1 cDNA was prepared and cloned into the pRAT vector. Incorporation of SeMet was confirmed by mass spectro-

Table 1
Data-collection and processing statistics.

Values for the highest resolution shell are shown in parentheses.

|	Cgl-1 (native)	Cgl-1 (SeMet)			Cgl-2
		Peak	Inflection	Remote	
Wavelength (Å)	0.9049	0.9785	0.9788	0.9636	1.0043
Space group	$C222_1$	$C222_1$			$P2_1$
Unit-cell parameters (Å, °)	a = 67.7, b = 80.9, c = 121.2	a = 67.5, b = 80.5, c = 121.0			a = 44.4, b = 119.8 c = 59.1, β = 99.3
Resolution (Å)	21–1.85	25–2.0			16–1.60
R_{sym} (%)	4.3 (4.9)	4.4 (6.1)	4.4 (6.2)	4.2 (5.8)	3.9 (7.1)
R_{anom} (%)		5.4 (6.6)	2.7 (3.8)	3.7 (4.7)	
Completeness (%)	99.7 (100)	100 (100)	100 (100)	100 (100)	98.1 (97.1)
Measured reflections	118291	132858	132888	132136	337900
Unique reflections	28719	22729	22712	22641	78609
Average $I/\sigma(I)$	10.5 (11.7)	9.5 (9.3)	10 (9.1)	10.5 (9.6)	10.7 (9.4)
Multiplicity	4.1	5.8	5.9	5.8	4.3
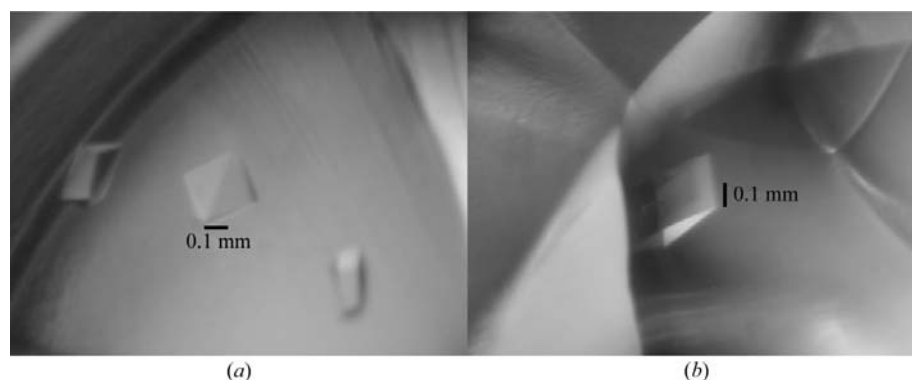
Figure 2
Crystals of recombinant Cgl-1 (*a*) and Cgl-2 (*b*) obtained using the sitting-drop vapour-diffusion method.

metry. The measured molecular weight reflected essentially 100% mutation-driven SeMet incorporation. The purified protein showed no change in stability relative to its parental form. Experimental measurements were readily performed using a tunable synchrotron-radiation source. Selenium sites in SeMet-substituted Cgl-1 were located by analyzing our scaled MAD data with *CNS* (Brünger *et al.*, 1998). Using data to 2.0 Å, six of the ten expected selenium sites were placed by *CNS* and generated an initial electron-density map with an overall figure of merit of 0.51. Subsequent density modification improved the figure of merit to 0.62. At present, model building and refinement are in progress for Cgl-1. This result will also enable the structural determination of native Cgl-2 by Patterson search techniques using the Cgl-1 structure as a starting model in order to accomplish the envisaged comparative analysis.

References

Boulianne, R. P., Liu, Y., Aebi, M., Lu, B. C. & Kües, U. (2000). *Microbiology*, **146**, 1841–1853.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Collaborative Computational Project, Number 4 (1994) *Acta Cryst.* D**50**, 760–763.

Cooper, D. N. W. (2002). *Biochim. Biophys. Acta*, **1572**, 209–231.

Cooper, D. N. W., Boulianne, R. P., Charlton, S., Farrell, E. M., Sucher, A. & Lu, B. C. (1997). *J. Biol. Chem.* **272**, 1514–1521.

Di Donato, A., de Nigris, M., Russo, N., Di Biase, S. & D'Alessio, G. (1993). *Anal. Biochem.* **212**, 291–293.

Gabius, H.-J. (1997). *Eur. J. Biochem.* **243**, 543–576.

Gabius, H.-J. (1998). *Pharmaceut. Res.* **15**, 23–30.

Gabius, H.-J. (2001). *Biochimie*, **83**, 659–666.

Gassner, N. C. & Matthews, B. W. (1999). *Acta Cryst.* D**55**, 1967–1970.

Hirabayashi, J. (1997). *Trends Glycosci. Glycotechnol.* **9**, 1–180.

Hirabayashi, J., Hashidate, T., Arata, Y., Nishi, N., Nakamura, T., Hirashima, M., Urashima, T., Oka, T., Futai, M., Müller, W. E. G., Yagy, F. & Kasai, K. (2002). *Biochim. Biophys. Acta*, **1572**, 232–254.

Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992). *Comput. Appl. Biosci.* **8**, 275–282.

Kaltner, H. & Stierstorfer, B. (1998). *Acta Anat.* **161**, 162–179.

Kopitz, J., André, S., von Reitzenstein, C., Versluis, K., Kaltner, H., Pieters, R. J., Wasano, K., Kuwabara, I., Liu, F.-T., Cantz, M., Heck, A. J. R. & Gabius, H.-J. (2003). *Oncogene*, **22**, 6277–6288.

Leslie, A. G. W. (1997). *MOSFLM User's Guide: MOSFLM* v.5.50. MRC Laboratory of Molecular Biology, Cambridge, UK.

Lis, H. & Sharon, N. (1998). *Chem. Rev.* **98**, 637–674.

Liu, F.-T., Patterson, R. J. & Wang, J. L. (2002). *Biochim. Biophys. Acta*, **1572**, 263–273.

Loris, R. (2002). *Biochim. Biophys. Acta*, **1572**, 198–208.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Peränen, J., Rikkonen, M., Hyvönen, M. & Kääriäinen, L. (1996). *Anal. Biochem.* **236**, 371–373.

Rabinovich, G. A., Rubinstein, N. & Toscano, M. A. (2002). *Biochim. Biophys. Acta*, **1572**, 274–284.

Rüdiger, H., Siebert, H.-C., Solís, D., Jiménez-Barbero, J., Romero, A., von der Lieth, C.-W., Díaz-Mauriño, T. & Gabius, H.-J. (2000). *Curr. Med. Chem.* **7**, 389–416.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, edited by N. Ford & C. Nolan, pp. A.1–A.13. Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press.

Solís, D., Jiménez-Barbero, J., Kaltner, H., Romero, A., Siebert, H.-C., von der Lieth, C.-W. & Gabius, H.-J. (2001). *Cells Tissues Organs*, **168**, 5–23.

Yagi, F., Miyamoto, M., Abe, T., Minami, Y., Tadera, K. & Goldstein, I. J. (1997). *Glycoconj. J.* **14**, 281–288.