Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Purification, crystallization and preliminary X-ray diffraction analysis of the carbohydrate-binding domain of flocculin, a cell-adhesion molecule from *Saccharomyces carlsbergensis*

Magnus Groes,[a,b] Kaare Teilum,[b] Kjeld Olesen,[c] Flemming M. Poulsen[b] and Anette Henriksen[a]*

[a]Carlsberg Laboratory, Department of Chemistry, Gamle Carlsberg Vej 10, DK-2500 Valby, Denmark, [b]Institute of Molecular Biology, University of Copenhagen, Øster Farimagsgade 2A, DK-1353 København K, Denmark, and [c]Carlsberg Research Laboratory, Gamle Carlsberg Vej 10, DK-2500 Valby, Denmark

Correspondence e-mail: anette@crc.dk

Received 24 June 2002
Accepted 27 August 2002

The recombinant carbohydrate-binding domain of the cell-surface lectin flocculin from brewer's yeast has been identified, purified and crystallized. The expression of the protein is associated with the nutritional state of the yeast. $P2_12_12_1$ crystals with unit-cell parameters a = 36.5, b = 59.7, c = 83.1 Å were obtained in hanging drops at 295 K using 25%(w/v) PEG 4000, 0.05 M KH_2PO_4 as precipitant. X-ray diffraction data have been obtained to 2.6 Å. The asymmetric unit contains one molecule and has a solvent content of 32%. An isomorphous $PtCl_4^{2-}$ derivative has been obtained.

1. Introduction

Lectins are a structurally diverse group of proteins capable of binding carbohydrates. They mediate a variety of biological processes, such as cell–cell and host–pathogen interactions, and are found in most organisms, ranging from viruses and bacteria to plants and animals. Mechanisms for sugar recognition have evolved independently in diverse protein frameworks, but some general features are observed. The binding sites are typically placed on the surface of the proteins, in contrast to many sugar-processing enzymes, where the substrate binding often occurs in a deep cleft. Furthermore, the lectin binding sites generally have relatively low affinity but considerable specificity for saccharides.

In the past few years, the structures of lectins have been extensively studied and more than a hundred three-dimensional structures are now available. The plant lectins in general, and legume lectins in particular, are the most studied and best characterized fractions of the known lectins. However, the structural flocculins from brewer's yeast contain lectin domains which are neither structurally characterized nor categorized in any known lectin family. It has been hypothesized that flocculins are structurally closely related to legume lectins since they share some key features (Teunissen & Steensma, 1995), having sizes around 25–30 kDa and binding calcium. Furthermore, flocculins contain many residues that promote formation of the antiparallel β-pleated sheets found in legume lectins. However, the flocculins do not have sequence homology to any known lectins and are therefore more likely to exhibit a new lectin fold.

Yeast flocculation, the adhesion of the cells in clumps resulting in their sedimentation from the medium, has been reported to be affected by many genes including the dominant gene *FLO1* (Watari *et al.*, 1994). *FLO1* encodes a cell-wall protein conferring flocculation ability to the cells in the absence of mannose (Bidard *et al.*, 1995). A *FLO1* homologue, *Lg-FLO1*, has been shown to encode a mannose/glucose-specific lectin, giving rise to a mannose/glucose-inhibited flocculation phenotype instead of the mannose-inhibited flocculation phenotype observed in strains carrying the *FLO1* gene (Kobayashi *et al.*, 1998). The open reading frame of the *FLO1* gene encodes a protein of 1537 amino acids composed of repeated sequences and includes an N-terminal sugar-binding domain of approximately 240 residues (Kobayashi *et al.*, 1998; Teunissen *et al.*, 1993; Watari *et al.*, 1994). Here, we report the construction of an expression plasmid for the lectin domain of Lg-Flo1p from *Saccharomyces carlsbergensis*, its crystallization and the preliminary X-ray diffraction analysis.

2. Materials and methods

2.1. Construction of an expression plasmid for the lectin domain of Lg-Flo1p

Based on the published DNA sequences of the N- and C-terminal parts of *Lg-FLO1* [GeneBank entries D89860 (1996) and AB003521 (1997)], two oligonucleotides [KOL97-B16, 5′-d(TTACGAATTCCTATTA-AGTTGTTGGTGTTTTGACAACAATG)-3′ and KOL98-03, 5′-d(AGCAAAGCTTAAA-AAAATGACAATTGCACACCA)-3′ (DNA Technology)] were designed in order to amplify a 1 kbp sequence from *S. carlsbergensis* strain CG2164. The amplified PCR product was sequenced from both ends and based on this new sequence information, two oligonucleotides [FLO1-CTERM, 5′-d(A-

AGGAAAAAAGCGGCCGCTTATTAAGTATGTTTTGAAGGATCAGG)-3′ and KOL99-13, 5′-d(CCGCTCGAGAAAAGAGAGGCTGAAGCTACACAAGCATGCCTGCCA)-3′ (DNA Technology)] were designed in order to amplify and subsequently clone the full-length lectin domain from brewer's yeast. The PCR product was digested with *Not*I and *Xho*I and ligated into the pPIC-9k expression vector (Invitrogen) in frame with the α-signal peptide of the vector, targeting the lectin domain for secretion to the growth medium. Since pPIC-9k contains two *Xho*I sites, a partial cleavage with *Xho*I was performed and the band with the highest MW was extracted from an agarose gel. Subsequently, this DNA was cleaved with *Not*I and again the band with the highest MW was extracted from an agarose gel. The ligated plasmid was transformed into *Escherichia coli* (DH5α) and a midi-preparation of the plasmid was cut with *Sal*I prior to the transformation into the *AOX1* locus of *Pichia pastoris* strain GS115.

2.2. Expression and purification of the lectin domain of Lg-Flo1p

The protein was expressed by fermentation in flasks. *P. pastoris* was grown in minimal medium according to the following procedure. 25 ml BMG (1% glycerol) (Invitrogen *Pichia* manual) was inoculated overnight at 303 K, followed by inoculation of 225 ml BMG with the 25 ml BMG culture for 24 h at 303 K to OD_{600} = 8. The cells were harvested by centrifugation (1000*g*, 10 min) and suspended in 1 l BMM (0.5% methanol) (Invitrogen *Pichia* manual) to OD_{600} = 2. They were then grown for 72 h at 303 K with 5 ml methanol (100%) added every 24th hour.

The lectin domain was purified from the growth medium after removing the cells by centrifugation. The suspension was concentrated from 1 l to 50 ml in a 350 ml Amicon cell at 355 kPa (YM10 filter) and dialyzed (molecular-weight cutoff 3500 Da) against water. At a conductivity of $\varepsilon < 2$ mS the sample was separated into five portions and each was purified by ion-exchange chromatography (6 ml Q-resource column; Pharmacia) at a flow rate of 2.5 ml min^{-1} at pH 8. The protein was eluted in 50 m*M* Tris buffer with a salt gradient starting at 0 m*M* NaCl and increasing by 4.5 m*M* ml^{-1}. Fractions were collected at around 150 m*M* NaCl and the total pool was concentrated to 5 ml by centrifugation in a centriprep YM10 at 3000*g*. The 5 ml was gel filtrated on a 100 ml G75 column at a flow rate of 1 ml min^{-1} in 50 m*M* NH_4HCO_3 buffer, followed by ion-exchange chromatography on a 6 ml Q-Resource column at a flow rate of 2.5 ml min^{-1} at pH 5.5. A salt gradient was run in 50 m*M* piperazine buffer, starting from 0 m*M* NaCl and increasing by 4.5 m*M* ml^{-1}. Fractions were collected at around 100 m*M* NaCl.

The *Lg-FLO1* gene-fragment product is heterogeneously glycosylated in the *P. pastoris* expression system; enzymatic deglycosylation was performed to minimize the heterogeneity of the sample. The protein was treated twice with 10 µl EndoH (New England Biolabs, 500 U µl^{-1}) and incubated at 298 K overnight, between which the protein was dialyzed against the buffer. The homogeneity and molecular mass of the sample were evaluated by SDS–PAGE and N-terminal amino-acid sequencing. Finally, the protein was gel filtrated (see above), concentrated and dissolved in 10 m*M* Tris pH 7 to a concentration of 8 mg ml^{-1}. The concentration was estimated from the A_{280}.

2.3. Mannose-binding assay

The binding of mannose to Lg-Flo1p was studied by fluorescence spectroscopy using an excitation wavelength of 275 nm, an emission wavelength of 322 nm, a scan speed of 50 nm min^{-1}, an excitation slit width of 10 nm and an emission slit width of 5 nm. The intensity of the emission peak at 322 nm is increased when mannose is bound to the protein. The initial sample volume was 2 ml containing 5 µ*M* Lg-Flo1p and buffer [0.08% Brij 35 (Sigma), 50 m*M* acetate pH 5.0 and 1 m*M* $CaCl_2$]. This was titrated with 2–80 µl solution containing 5 µ*M* Lg-Flo1p, buffer and sugar. At each titration step the sample was mixed three times with a mixer and the fluorescence measured immediately. The duration of each step was about 3 min. The titration curve was fitted using the program package *GraFit 3.09b* for a single-site ligand-binding protein.

2.4. Crystallization, data collection and data processing

Crystal Screen I supplied by Hampton Research (California, USA) was used for initial screening of crystallization conditions using the hanging-drop vapour-diffusion method (Jancarik & Kim, 1991). The initial crystallization conditions were optimized by varying the PEG concentration, the protein concentration, the concentration of KH_2PO_4 and by addition of salts from Additive Screen I (Hampton Research). The crystals achieved under the optimized crystallization conditions could be transferred to a cryoprotectant consisting of 35%(*w*/*v*) PEG 4000, 0.05 *M* KH_2PO_4 and flash-frozen in liquid nitrogen for storage and data collection. Diffraction data from native flocculin crystals were collected at the MAXLAB II synchrotron (University of Lund, Sweden) on beamline 711. A K_2PtCl_4 derivative of flocculin was prepared by soaking the crystals in 10 m*M* K_2PtCl_4, 35%(*w*/*v*) PEG 4000 for 3 h. Data from the heavy-atom derivative were also collected at the MAXLAB II synchrotron beamline 711. In both cases, the diffraction data were collected as a series of discrete frames, with an oscillation width of 1.0° for the individual frames and a total oscillation width of 120° for the native data and of 143° for the K_2PtCl_4 data. Data indexing, integration and scaling were performed using *MOSFLM* (Leslie, 1999) and *SCALA* (Collaborative Computational Project, Number 4, 1994). Heavy-atom positions and preliminary phases were calculated in *SOLVE* (Terwilliger & Berendzen, 1999) and *CNS* (Brünger *et al.*, 1998).

3. Results and discussion

3.1. Cloning and expression of Lg-FLO1

Using the published DNA sequences coding for the N- and C-terminal parts of Lg-Flo1p (Kobayashi *et al.*, 1998), we identified the previously unknown 174 base pairs of the *Lg-FLO1* gene sequence. The region constituting the lectin domain of Lg-Flo1p were defined by the characteristic flanking sequences: a signal sequence at the N-terminal end and a cell-wall-spanning repeated sequence at the C-terminal end. Thus, based on the full-length sequence of *Lg-FLO1*, we could amplify the region hypothesized to constitute the lectin domain of Lg-FLO1p and clone it into the general *P. pastoris* expression vector pPIC-9k. From cultures of *P. pastoris* transformed with the resulting plasmid, we were able to purify 20 mg protein per litre of medium. The DNA sequence of the cloned *Lg-FLO1* gene fragment was determined and was in accordance with the published sequence except from the substitution of a thymine with a cytosine, giving rise to an alanine instead of a valine at position 179 (Fig. 1). N-terminal protein sequencing revealed that the EAEA part of the introduced secretion signal was not processed by *P. pastoris* (Fig. 1). Constructs made to prevent the EAEA sequence in the product gave low expression levels and were abandoned. The final recombinant Lg-Flo1p consisted of 224 amino acids, including four cysteine residues

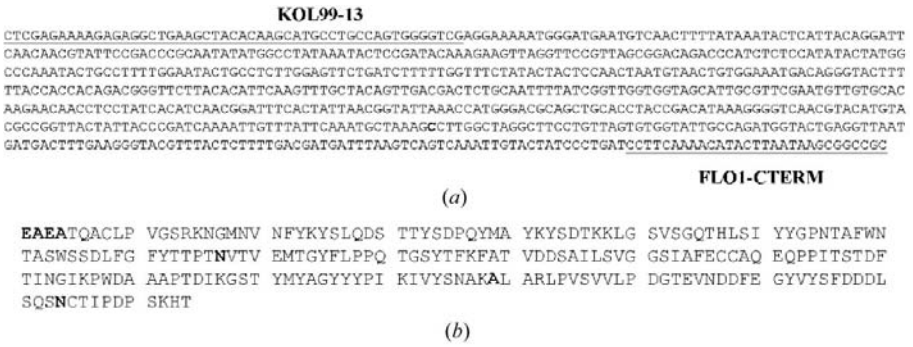

Figure 1
(*a*) The nucleotide sequence of the cloned *Lg-FLO1* gene fragment, with primers underlined. (*b*) The amino-acid sequence of the purified lectin-like domain from flocculin. The sequence in bold is the *P. pastoris* secretion signal still present in the final gene product.

Table 1
Data-collection statistics.

Values in parentheses refer to the highest resolution bin.

Crystal	Native	K_2PtCl_4
Beamline	MAX-LAB BL-711	MAX-LAB BL-711
Wavelength (Å)	0.968	0.968
Temperature (K)	100	100
Unit-cell parameters (Å)	a = 36.5, b = 59.7, c = 83.1	a = 36.1, b = 59.9, c = 82.3
Solvent content (%)	32	31
Resolution range	9.99–2.60	9.99–2.52
Total No. of reflections	26211	31822
No. of unique reflections	5607	5666
Completeness	97.7 (99.3)	97.9 (89.1)
R_{merge}†	0.091 (0.175)	0.049 (0.212)
$I/\sigma(I)$	6.7 (3.9)	5.9 (3.3)

† $R_{merge} = \sum|I - \langle I\rangle|/\sum I$, where I is the intensity of an individual observation and $\langle I\rangle$ is the average intensity from multiple observations.

and two potential N-glycosylation sites. Owing to glycosylation, four bands with 3.7 < pI < 4.3 were observed with IEF, but after deglycosylation with EndoH the resulting 25 kDa protein ran as a single species on an IEF gel, with a pI of 4.3.

3.2. Mannose binding

To confirm the functional integrity of Lg-Flo1p, a fluorescence assay was performed. The K_D of Lg-Flo1p was determined to be 0.77 (0.028) mM by non-linear regression of the intensity changes by titration of mannose.

3.3. Crystallization

The optimized crystallization conditions consisted of a hanging drop made from equal volumes of protein (6 mg l^{-1}) and reservoir solution, equilibrated against reservoir solution consisting of 25%(*w*/*v*) PEG 4000, 0.05 *M* KH_2PO_4 at 295 K. Under these conditions, it was necessary to microseed the drops to achieve crystallization. In most cases the crystals grew as multinucleated bundles of needles, but occasionally single crystals were obtained. Of these, the largest crystals were rod-shaped with dimensions of 0.05 × 0.05 × 0.2 mm.

3.4. X-ray diffraction data collection and preliminary analysis

The Lg-FLO1p lectin-domain crystals belong to space group $P2_12_12_1$, with unit-cell parameters as listed in Table 1 and a maximum resolution of 2.6 Å. Initial unit-cell and space-group assignment used the *MOSFLM* (Leslie, 1999) autoindexing routine. Final statistics of the data processing are given in Table 1. The unit-cell parameters correspond to a solvent content of 32%, assuming one molecule per asymmetric unit.

We thank the EU ARI programme for financial support for the work at the MAXLAB II synchrotron (Lund). DANSYNC is acknowledged for financial support. The beamline scientist Yngve Cerenius (MAXLAB) is thanked for technical assistance and support.

References

Bidard, F., Bony, M., Blondin, B., Dequin, S. & Barre, P. (1995). *Yeast*, **11**, 809–822.
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.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Kobayashi, O., Hayashi, N., Kuroki, R. & Sone, H. (1998). *J. Bacteriol.* **180**, 6503–6510.
Leslie, A. (1999). *MOSFLM.* MRC Laboratory of Molecular Biology, Cambridge, England.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Teunissen, A. W., Holub, E., van der Hucht, J., van den Berg, J. A. & Steensma, H. Y. (1993). *Yeast*, **9**, 423–427.
Teunissen, A. W. & Steensma, H. Y. (1995). *Yeast*, **11**, 1001–1013.
Watari, J., Takata, Y., Ogawa, M., Sahara, H., Koshino, S., Onnela, M. L., Airaksinen, U., Jaatinen, R., Penttila, M. & Keranen, S. (1994). *Yeast*, **10**, 211–225.