Acta Cryst. (1997). D53, 469–471

Homology between jacalin and artocarpin from jackfruit (*Artocarpus integrifolia*) seeds. Partial sequence and preliminary crystallographic studies of artocarpin

STEPHEN SURESH, P. GEETHA RANI, J. VENKATESH PRATAP, R. SANKARANARAYANAN, A. SUROLIA AND M. VIJAYAN* *at Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India. E-mail: mv@mbu.iisc.ernet.in*

(Received 11 October 1996; accepted 20 January 1996)

Abstract

Jacalin and artocarpin, the two lectins from jackfruit (*Artocarpus integrifolia*) seeds, have different physicochemical properties and carbohydrate-binding specificities. However, comparison of the partial amino-acid sequence of artocarpin with the known sequence of jacalin indicates close to 50% sequence identity. Artocarpin crystallizes in two forms, both monoclinic $P2_1$, with one and two tetramic molecules, respectively, in the asymmetric units of form I ($a = 69.9$, $b = 73.7$, $c = 60.6$ Å and $\beta = 95.1°$) and form II ($a = 87.6$, $b = 72.2$, $c = 92.6$ Å and $\beta = 101.1°$). Both the crystal structures have been solved by the molecular replacement method using the known structure of jacalin as the search model and one of them partially refined, confirming that the two lectins are indeed homologous.

1. Introduction

Lectins are well known for their ability to specifically recognize cell-surface carbohydrates and this property has led to their considerable importance in biological research and biomedical applications (Liener, Sharon & Goldstein, 1986; Sharon & Lis, 1989). Lectins from plant, animal, bacterial and viral sources have been studied in great detail, with those from plants receiving the maximum attention (Rini, 1995). Among plant lectins, those from leguminous plants share a common tertiary fold although they exhibit widely different quaternary structures and sugar specificities (Delbaere *et al.*, 1990; Shaanan, Lis & Sharon, 1991; Banerjee *et al.*, 1994). The lectin component of the plant toxin ricin, cereal lectins represented by wheatgerm agglutinin and lectins belonging to the *Amarillidaceae* family represented by snowdrop lectin, have been shown to have distinctly different charateristic tertiary folds (Rini, 1995; Drikamer, 1995). As part of a preogramme of structural studies on lectins (Banerjee *et al.*, 1994, 1996; Sankaranarayan *et al.*, 1993), we have recently shown that jacalin, a lectin from the jackfruit (*Artocarpus integrifolia*) seeds, has yet another lectin fold and a novel carbohydrate binding site (Sankaranarayan *et al.*, 1996). The present report is concerned with the second lectin, artocarpin, isolated from the jackfruit seeds, which among other things mediates the activation of T-cell dependent B-cell maturation (de Miranda-Santos, Mengel, Bunn-moreno & Campos-Neto, 1991). The physicochemical properties and the carbohydrate-binding specificity of this lectin have been studied in considerable detail in this laboratory (Misquith, Rani, & Surolia, 1994). Unlike other mannose-binding lectins, artocarpin displays exquisite specificity to the 1,6 fucosylated, xylose bissected *N*-linked core oligosaccharide chain such as the one found in horseradish peroxidase.

Artocarpin, M_r 65 000, differs from jacalin, M_r 66 000, in many important respects. The subunit of the former is made up of a single polypeptide chain, while that of the latter contains two, although both are homotetrameric. Jacalin with a high degree of microheterogeneity in sequence is a glycoprotein while artocarpin consisting of four isolectins is not. Artocarpin is mannose specific at the monosaccharide level and has the strongest affinity for the xylose containing heptasaccharide from horseradish peroxidase (Misquith, Rani & Surolia, 1994). Jacalin, on the other hand, is galactose specific and has the highest affinity for the T-antigenic disaccharide Galβ1–3 GalNAc. We show here that despite these differences, the two lectins are homologous.

2. Materials and methods

2.1. *Purification and sequencing*

Artocarpin was purified using mannose–sepharose affinity chromatography (Misquith *et al.*, 1994). The eluted protein was extensively dialyzed against water and lyophilized. The lyophilized protein was treated with 8 *M* guanidium hydrochloride (GdmCl) in 0.1 *M* ammonium bicarbonate (pH 8.0) for 1 h at 298 K and then diluted to 2 *M* GdmCl with 0.1 *M* ammonium bicarbonate. Tryptic cleavage was then carried out with tosyl phenyl chloromethyl ketone–trypsin at an enzyme: substrate ratio of 1:50(*w*/*w*) at 310 K for 6 h. The digestion was terminated by acidification with 25% trifluoroacetic acid (TFA) to pH = 2.5. The soluble fraction was separated by centrifugation prior to their purification on C_2–C_{18} by reverse-phase fast protein liquid chromatography (RP-FPLC) using a linear gradient of acetonitrile 5–60% in 0.1% TFA spanning 40 min. A total of 12 peaks in RP-FPLC were obtained. The peaks eluting at 29, 36, 39 and 44% of acetonitrile designated as T1, T2, T3 and T4 gave unambiguous amino-acid sequences. Iodobenzoic acid (IBA) cleavage was performed (Fontana, Dalzoppo, Grandi & Zambonin, 1983). The digest was electrophoresed on tricine gel (Schagger & Jagow, 1987). Tricine gel displayed two bands in addition to the band corresponding to the uncleaved protein. The slower and faster moving bands are designated as IBA1 and IBA2, respectively. IBA1 and IBA2 were then transferred onto polyvinyl difluoride membrane. N-terminal analysis of the native artocarpin and the peptides were carried out on Shimadzu Automated gas phase sequencer (model PSQ-1) equipped with an online C-R4A120A Chromatopac Shimadzu Phenylthiohydrantoin (PTH) Analyzer. The initial yields were 60–90% and the repetitive yields were above 93%.

2.2. *Crystallization*

Crystallization experiments were performed using the vapour-diffusion method. In a typical experiment, 10 µl of 10 mg ml^{-1} protein (6.3 µ*M*) in 0.1 *M* phosphate buffer at pH 7.4, containing 0.1 *M* NaCl and 0.025%(*w*/*v*) sodium azide, was mixed with 2 µl of 20%(*w*/*v*) PEG 1450, and the whole solution was equilibrated at 293 K with a reservoir solution of 40%(*w*/*v*)

© 1997 International Union of Crystallography
Printed in Great Britain – all rights reserved

PEG 4000 in the same buffer. The crystals grew to their maximum size (0.5 × 0.5 × 0.2 mm) in 10–15 d. Two distinct crystal forms (referred to as forms I and II) were obtained under similar conditions.

2.3. *X-ray crystallographic studies*

The space group and the unit-cell dimensions of the crystals were determined using precession photography. The solvent contents were determined by the method of Matthews (Matthews, 1968). Diffraction data from crystals of form I were collected on a Siemens–Nicolet area-detector system mounted on a GX20 Marconi Avionics rotating-anode X-ray generator, operating at 40 kV and 40 mA. The raw data were reduced using the *XENGEN* software (Howard *et al.*, 1987). The data for the form II crystals were obtained on a MAR imaging plate mounted on a Rigaku rotating-anode X-ray generator, operating at 40 kV and 50 mA. The *XDS* data-reduction software package (Kabsch, 1988) was used to reduce the raw data. Molecular replacement calculations were carried out using *AMoRe* (Navaza, 1994) while the *X-PLOR* package (Brünger, Kuriyan & Karplus, 1987) was used for initial structure refinement.

3. Results and discussion

The amino-acid sequences of the six peptide fragments of artocarpin, accounting for about 60% of the length of the polypeptide chain, are given in Fig. 1. The stretches of the jacalin sequence, which aligned well with these sequences, are also shown in the figure. The sequences account for 87 amino-acid residues and 46 of them are identical in the two proteins. Thus the partial sequences of the two proteins exhibit nearly 50% identity, which strongly indicate that the two are homologous.

The crystal data of the two forms are given in Table 1. The solvent content of form I is 49% for one tetramer in the asymmetric unit while it is 45% for form II for two tetramers in the asymmetric unit. Intensity data were collected to a resolution of about 2.4 Å from the crystals of forms I and II with merging R values of 0.090 and 0.089 and completion of 87 and 96%, respectively. In the case of form I, 74% of the measured reflections upto a resolution of 2.6 Å had $F > 2\sigma(F)$ while it was 32% for reflections beyond 2.6 Å resolution. The corresponding values in the case of form II were higher at 83 and 61%. Reflections in the resolution range 15.0–3.5 Å were used in both the cases for molecular replacement calculations with the jacalin tetramer (Sankaranarayanan *et al.*, 1996) as the search model. One tetramer in form I (correlation factor 0.434 and R factor 0.422) and two tetramers in form II (correlation factor 0.522 and R factor 0.460) could be readily located from these calculations. Calculations using a jacalin dimer as the search model confirmed these solutions. Packing considerations indicated the solutions to be acceptable. Rigid-body refinement of the tetramers placed in the unit cell, with the jacalin sequence replaced by polyalanine, was then carried out using 10.0–3.0 Å data. Subsequently, the polyalanine models were further refined using data in the same resolution shell which yielded R factors of 0.374 and 0.381, respectively, for forms I and II. Electron-density maps were calculated at this stage and 50 of the 87 side chains known from the partial sequence were progressively fitted into the electron-density map corresponding to form I. This partial model refined to an R value of 0.310 and R free of 0.384 for 10 572 observed reflections [$F > 2\sigma(F)$] in the 10.0–3.0 Å resolution range, indicating the essential correctness of the structure.

The high level of sequence identity in the partial sequences of jacalin and artocarpin, the successful molecular replacement solution of the two crystal structures using jacalin as the search model and the ready refinement of the partial structure of one of the crystal forms to an acceptable level, clearly show that the two lectins are homologous. The determination of the complete sequence of artocarpin and the refinement of its structure in the two crystal forms, are in progress. The work, when completed, is expected to reveal the structural basis of the different physicochemical and carbohydrate-binding properties of artocarpin and jacalin, and provide useful insights into the evolutionary relationship between the two lectins.

Table 1. *Preliminary X-ray data of the two crystal forms of artocarpin*

Precipitant	Space group	a (Å)	b (Å)	c (Å)	β (°)	Solvent content (%)	Z
PEG 1450	$P2_1$	69.9	73.7	60.6	95.1	49	2
PEG 1450	$P2_1$	87.6	72.2	92.6	101.1	45	4

T1 : **F P** D **E** F L E I **V S G Y**

60 71

Jacalin : **F P** S **E** Y I M E **V S G Y**

T2 : E **A I G** S **F** S **V I Y** G Q T **G**

23 36

Jacalin : T **A I G** D **F** Q **V I Y** L L N **G**

T3 : T F **G P Y G D E E G T Y** F Y

92 105

Jacalin : N I **G P Y G D E E G T Y** L N

T4 : **N L P I E** D A I G **V** H M A L

105 118

Jacalin : **N L P I E** N G L I **V** G F K G

IBA1 : **D** E **G** S Y **T G I R** Q **I** E **L** L **Y** K T

5 21

Jacalin : **D** D **G** A F **T G I R** E **I** N **L** S **Y** N K

IBA2 : D D P G Y N T P **V V** G E **L T F K**

72 87

Jacalin : T G N V S G Y V **V V** R S **L T F K**

Fig. 1. Sequences of the peptide fragments of artocarpin. The corresponding stretches in the jacalin sequence, obtained through manual alignment are also indicated along with residue numbers. Identical residues are highlighted.

The work was supported by the Department of Science and Technology (DST), Government of India. The sequencing was carried out on the Peptide and Protein sequencing facility, supported by the Department of Biotechnology (DBT), X-ray data were collected using the Area Detector/Image Plate Facility supported by the DST and the DBT, and the computations were carried out at the Supercomputer Education and Research Centre and the DBT supported Graphics facility at the Institute.

References

Banerjee, R., Das, K., Ravishankar, R., Suguna, K., Surolia, A. & Vijayan, M. (1996). *J. Mol. Biol.* **259**, 281–296.

Banerjee, R., Mande, S. C., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S. K., Suguna, K., Surolia, A. & Vijayan, M. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 227–231.

Brünger, A. T., Kuriyan, J., Karplus, M. (1987). *Science*, **235**, 458–460.

Delbaere, L. T. J., Vandonselaar, M., Prasad, L., Quail, J. W., Pearlstone, J. R., Carpenter, M. R., Smillie, L. B., Nikrad, P. V., Spohr, U. & Lemieux, R. U. (1990). *Can. J. Chem.* **68**, 1116–1121.

Drickamer, K. (1995). *Nature Struct. Biol.* **2**, 437–439.

Fontana, A., Dalzoppo, D., Grandi, C. & Zambonin, M. (1983). *Methods Enzymol.* **91**, 311–318.

Howard, A. J., Gilliland, G. L., Finzel, B. C., Poulos, T. L., Ohlendorf, D. H. & Salemme, F. R. (1987). *J. Appl. Cryst.* **20**, 383–387.

Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 916–924.

Liener, I. E., Sharon, N. & Goldstein, I. J. (1986). Editors. *The Lectins: Properties, Functions and Applications in Biology and Medicine*, p. 600. Orlando: Academic Press.

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

de Miranda-Santos, I. K. F., Mengel, J. O. Jr, Bunn-moreno, M. M. & Campos-Neto, A. (1991). *J. Immunol. Methods*, **140**, 197–203.

Misquith, S., Rani, P. G. & Surolia, A. (1994). *J. Biol. Chem.* **269**, 30393–30401.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Rini, J. M. (1995). *Annu. Rev. Biophys. Biomol. Struct.* **24**, 551–577.

Sankaranarayanan, R., Puri, K. D., Ganesh, V., Banerjee, R., Surolia, A & Vijayan, M. (1993). *J. Mol. Biol.* **229**, 558–560.

Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A & Vijayan, M. (1996). *Nature Struct. Biol.* **3**, 596–603.

Schagger, H. & Jagow, G. V. (1987). *Analyt. Biochemistry*, **166**, 368–379.

Shaanan Lis, L. & Sharon, N. (1991). *Science*, **254**, 862–866.

Sharon, N. & Lis, L. (1989). *Lectins*, pp. 69–90. London: Chapman and Hall.