

Byung-Cheon Jeong,
Kyoung Shin Yoo, Kwang Wook
Jung, Jeong Sheop Shin and
Hyun Kyu Song*

School of Life Sciences and Biotechnology,
Korea University, Anam-dong, Seoungbuk-gu,
Seoul 136-701, Republic of Korea

Correspondence e-mail: hksong@korea.ac.kr

Received 24 June 2008
Accepted 5 August 2008

Purification, crystallization and preliminary X-ray diffraction analysis of a cystathionine β -synthase domain-containing protein, CDCP2, from *Arabidopsis thaliana*

Cystathionine β -synthase domain-containing protein 2 (CDCP2) from *Arabidopsis thaliana* has been overexpressed and purified to homogeneity. As an initial step towards three-dimensional structure determination, crystals of recombinant CDCP2 protein have been obtained using polyethylene glycol 8000 as a precipitant. The crystals diffracted to 2.4 Å resolution using synchrotron radiation and belonged to the trigonal space group $P3_121$ or $P3_221$, with unit-cell parameters $a = b = 56.360$, $c = 82.596$ Å, $\alpha = \beta = 90$, $\gamma = 120^\circ$. The asymmetric unit contains one CDCP2 molecule and the solvent content is approximately 41%.

1. Introduction

The cystathionine β -synthase (CBS) domain is an evolutionarily conserved protein domain and is defined by sequence motifs that have been identified in various proteins from all three kingdoms of life (Ignoul & Eggermont, 2005). It usually exists as tandem repeats and is found in cytosolic and membrane proteins that perform various functions. Tandem pairs of CBS domains can act as binding modules for adenosine derivatives and may regulate the activity of attached enzymes or other domains (Ignoul & Eggermont, 2005). Therefore, CBS domains may act as sensors of cellular energy status by being activated by AMP and inhibited by ATP (Hardie *et al.*, 2006). Proteins containing CBS domains have very diverse functions, ranging from metabolic enzymes and transcriptional regulators to ion channels and transporters. For example, pairs of CBS domains exist in AMP-activated protein kinase (AMPK; Townley & Shapiro, 2007), inosine-5-monophosphate dehydrogenase (IMPDH; Sintchak *et al.*, 1996) and chloride channel (CLC; Estevez *et al.*, 2004; Meyer *et al.*, 2007). Although their precise function remains to be elucidated, the physiological importance of CBS domains is emphasized by the observation that point mutations in CBS domains can seriously cripple the specific protein function and are responsible for several human hereditary diseases including homocystinuria (cystathionine β -synthase), Wolff–Parkinson–White syndrome (γ 2 subunit of AMP-activated protein kinase), retinitis pigmentosa (IMP dehydrogenase 1), congenital myotonia, idiopathic generalized epilepsy, hypercalciuric nephrolithiasis and classic Bartter syndrome (CLC chloride-channel family members; Bateman, 1997; Kemp, 2004).

CBS domain-containing proteins have been well characterized in mammalian systems, but those from plants, which are easily detected by sequence analysis and are also quite abundant (68 CBS domains were found in *Arabidopsis thaliana*; <http://pfam.sanger.ac.uk/family?acc=PF00571>), have been very poorly studied to date. Several structures of CBS domains from mammals, archaea and eubacteria have been reported (Ignoul & Eggermont, 2005; Townley & Shapiro, 2007; Miller *et al.*, 2004; Rudolph *et al.*, 2007; Day *et al.*, 2007), but no structural information is available on CBS domains from plant sources. Two CBS domain-containing proteins (CDCP1 and CDCP2; At4G34120 and At4G36910, respectively) from *A. thaliana*, which contain only one CBS domain each, have recently been cloned (unpublished results), but the exact functions of these two CDCPs cannot be predicted as proteins containing CBS domains can have

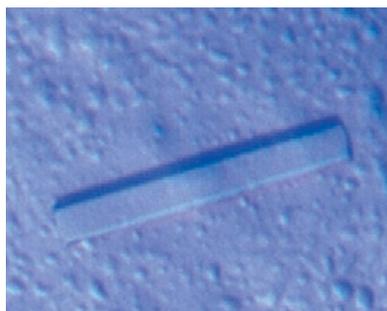


Table 1

Data-collection statistics.

Values in parentheses are for the highest resolution bin.

	Native	Selenomethionine derivative			
		Peak	Edge	High remote	Low remote
X-ray source	4A, PAL	NW12, PF			
Wavelength (Å)	1.00000	0.97923	0.97942	0.96000	0.98000
Space group	$P3_121$ or $P3_221$	$P3_121$ or $P3_221$			
Unit-cell parameters (Å)	$a = 55.81, c = 82.74$	$a = 56.36, c = 82.60$	$a = 56.41, c = 82.66$	$a = 56.44, c = 82.70$	$a = 56.49, c = 82.76$
Resolution (Å)	3.0 (3.11–3.0)	2.4 (2.49–2.40)	2.4 (2.49–2.40)	2.4 (2.49–2.40)	2.4 (2.49–2.40)
Unique reflections (no cutoff)	3242	6273	6270	6293	6265
Total reflections (no cutoff)	34806	63521	62425	62157	60947
Completeness (%)	99.9 (99.8)	99.4 (97.5)	99.4 (97.5)	99.4 (96.9)	98.8 (92.0)
Redundancy	10.7 (11.7)	10.1 (8.6)	10.0 (7.5)	9.9 (6.5)	9.7 (5.7)
$\langle I/\sigma(I) \rangle$	53.1 (12.79)	60.8 (4.25)	59.0 (3.11)	57.3 (2.48)	56.3 (2.00)
R_{merge} (%)	7.5 (41.7)	6.9 (41.2)	6.6 (53.2)	6.2 (55.2)	5.7 (59.6)

diverse functions. In order to understand the characteristics of plant CBS domain-containing proteins, structural study of the CDCPs from *A. thaliana* has been initiated. As the first step toward structure determination, CDCP2 has been overexpressed using an *Escherichia coli* expression system and purified to homogeneity. A suitable crystal for structure determination has been obtained and diffraction data have been collected from native and selenomethionyl CDCP2 crystals.

2. Materials and methods

2.1. Cloning, expression and purification

The gene encoding the conserved domain of CDCP2 (residues 72–236) was amplified by polymerase chain reaction (PCR) using the cDNA library of *A. thaliana*. The gene was flanked by *Bam*HI and *Xho*I restriction-enzyme sites and the fragments were ligated into a modified pET vector. The cloned vector contains a gene coding for an N-terminal GST tag followed by a TEV cleavage site. The integrity of the plasmids that were produced was verified by DNA sequencing. *E. coli* BL21 (DE3) cells were used to express the recombinant protein. Cells were grown in LB medium with kanamycin ($50 \mu\text{g ml}^{-1}$) to an $\text{OD}_{600\text{nm}}$ of 0.6; CDCP2 expression was then induced by addition of 1 mM IPTG at 303 K and maintained for 6 h. The culture was harvested by centrifugation for 20 min at $6000 \text{ rev min}^{-1}$. Harvested cells were resuspended in phosphate-

buffered saline (PBS) containing 1 mM PMSF and one tablet of protease-inhibitor cocktail (Roche) and then disrupted by ultrasonication. The lysate was collected by centrifugation for 1 h at 15 000g. The cell lysate was incubated with glutathione-linked resin which had been pre-equilibrated with PBS. The resin was washed extensively with PBS and the protein was eluted from the resin with elution buffer (50 mM Tris-HCl pH 8.0, 20 mM NaCl, 1 mM DTT and 10 mM reduced glutathione). CDCP2 was released from the fusion protein by digestion with 1:100(w:w) TEV protease at 295 K overnight. Only one extra glycine residue was attached to the N-terminus of CDCP2 after TEV cleavage. The released protein was diluted with buffer A [50 mM Tris-HCl pH 8.5, 1 mM DTT and 5% (v/v) glycerol] and loaded onto a HiTrap Q Fast Flow column equilibrated with buffer A. After the protein had been loaded, the column was washed with 25 ml buffer A. To elute the protein, a linear gradient was applied to buffer A containing 300 mM NaCl. CDCP2 was eluted at between 150 and 230 mM NaCl. Free GST contaminant was removed by passing the digested sample through a GST-Sepharose column. The pure fractions were concentrated using a 10 kDa cutoff filter (Amicon Ultra). The concentrated CDCP2 was further purified using a Superose 12 10/300 GL gel-filtration column in 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM DTT and 5% (v/v) glycerol. The concentration of purified CDCP2 was determined using the Bradford assay and the purity was checked by SDS-PAGE. Selenomethionine-substituted CDCP2 was expressed using B834 (DE3) cells and was purified in a similar manner to the wild-type protein.

2.2. Crystallization

The initial crystallization conditions for CDCP2 were obtained using DeCode Genetics Wizard II screen and Hampton Research Crystal Screen II with the hanging-drop vapour-diffusion method at room temperature (about 295 K) in 24-well VDX plates. Drops consisting of 2.0 μl purified CDCP2 (12.0 mg ml^{-1}) in 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM DTT, 5% (v/v) glycerol and 2.0 μl precipitant solution were equilibrated against 0.5 μl reservoir solution containing 100 mM phosphate-acetate pH 4.2, 200 mM NaCl and 20% (w/v) PEG 8000. Small crystals were obtained within 36 h and subsequently grew larger. In order to obtain a better crystal, grid screening was performed around the initial crystallization condition to find the optimal pH and precipitant concentration. To obtain the best crystal, oil (1:1 mixture of silicon oil and paraffin oil) was used over the reservoir solution (Blow *et al.*, 1994). The best crystal was obtained by placing 0.5 μl oil over 0.5 μl reservoir solution (Fig. 1). Crystals were transferred to a cryoprotectant solution containing



Figure 1

A trigonal crystal of *A. thaliana* CDCP2. Approximate dimensions are $0.1 \times 0.15 \times 0.5 \text{ mm}$.

reservoir buffer with 20% (v/v) glycerol and were then flash-frozen in a cold nitrogen stream at 100 K.

2.3. X-ray data collection

Diffraction data from native and selenomethionine-substituted crystals of CDCP2 were collected using the synchrotron-radiation sources beamline 4A at Pohang Accelerator Laboratory (PAL, Korea) and beamline NW12 at Photon Factory (PF, Japan), respectively; both data sets were recorded using an ADSC Quantum 210 CCD detector. All data sets were processed and integrated using *DENZO* and scaled using *SCALEPACK* from the *HKL-2000* program suite (Otwinowski & Minor, 1997). The native and selenomethionine-substituted crystals diffracted to 3.0 and 2.4 Å resolution, respectively. For multiwavelength anomalous dispersion (MAD) phasing, diffraction data were collected at four different wavelengths from a selenomethionyl CDCP2 crystal (Table 1).

3. Results and discussion

To obtain a well diffracting crystal of CDCP2, an additive screen, a detergent screen and a sodium chloride grid screen were performed. Initial attempts gave only tiny crystals of CDCP2; larger crystals were finally grown using oil over the reservoir. The oil acts as a barrier to vapour diffusion between the reservoir and the drop. 100% silicon oil or 100% paraffin oil were not effective as the oil barrier; using 100% silicon oil gave a result similar to that when no oil was used and using 100% paraffin oil only allowed an extremely limited amount of vapour diffusion. By mixing the two oils, the rate of vapour diffusion between the drop and the reservoir can be controlled. Al's oil is a 1:1 mixture of silicon oil and paraffin oil. The rate of vapour diffusion is also a function of the thickness of the oil layer over the reservoir. In the case of CDCP2 crystals, increasing the thickness of the oil layer over the reservoir seemed to produce proportionally larger sized crystals and the best crystal was obtained using 0.5 µl of Al's oil over 0.5 µl reservoir solution.

Table 1 summarizes the statistics of the synchrotron data collection. The native data consisted of 34 806 measurements of 3242 unique reflections, with an R_{merge} (on intensity) of 7.5%. Processing results showed that the CDCP2 crystal belonged to the trigonal crystal system. There was a clear systematic absence along a threefold axis. The space group is $P3_121$ or its enantiomorphic pair $P3_221$, with unit-cell parameters $a = b = 55.81$, $c = 82.74$ Å, $\alpha = \beta = 90$, $\gamma = 120^\circ$.

The asymmetric unit contained one CDCP2 molecule with a corresponding V_M of $2.08 \text{ \AA}^3 \text{ Da}^{-1}$ and a solvent content of approximately 41%. These values are within the frequently observed range for protein crystals (Matthews, 1968). For MAD experiments, energy scanning for Se atoms was performed using the selenomethionyl CDCP2 crystal and four wavelengths including a high-energy remote and a low-energy remote were chosen from the energy-scanning results. Although the crystallographic parameters are quite similar to those of the native data set, the diffraction limits and data quality for the selenomethionyl-derivatized crystal were better than those for the native crystal (Table 1). Subsequent experiments for MAD phasing are under way.

We thank the staff at beamline 4A of the Pohang Accelerator Laboratory (Pohang, Korea) and beamline NW12 of Photon Factory (Tsukuba, Japan) for help with data collection. This work was supported by grants from the Plant Signaling Network Research Center, Korea Science and Engineering Foundation and BioGreen21 program, Rural Development Administration, Korea. This research was also supported by a grant (CG3112 to JSS) from the Crop Functional Genomics Center of the 21st Century Frontier Research Program.

References

- Bateman, A. (1997). *Trends Biochem. Sci.* **22**, 12–13.
 Blow, D. M., Chayen, N. E., Lloyd, L. F. & Saridakis, E. (1994). *Protein Sci.* **3**, 1638–1643.
 Day, P., Sharff, A., Parra, L., Cleasby, A., Williams, M., Hörer, S., Nar, H., Redemann, N., Tickle, I. & Yon, J. (2007). *Acta Cryst.* **D63**, 587–596.
 Estevez, R., Pusch, M., Ferrer-Costa, C., Orozco, M. & Jentsch, T. J. (2004). *J. Physiol.* **557**, 363–378.
 Hardie, D. G., Hawley, S. A. & Scott, J. W. (2006). *J. Physiol.* **574**, 7–15.
 Ignoul, S. & Eggermont, J. (2005). *Am. J. Physiol. Cell Physiol.* **289**, C1369–C1378.
 Kemp, B. E. (2004). *J. Clin. Invest.* **113**, 182–184.
 Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
 Meyer, S., Savaresi, S., Forster, I. C. & Dutzler, R. (2007). *Nature Struct. Mol. Biol.* **14**, 60–67.
 Miller, M. D. *et al.* (2004). *Proteins*, **57**, 213–217.
 Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 301–326.
 Rudolph, M. J., Amodeo, G. A., Iram, S. H., Hong, S. P., Pirino, G., Carlson, M. & Tong, L. (2007). *Structure*, **15**, 65–74.
 Sintchak, M. D., Fleming, M. A., Futer, O., Raybuck, S. A., Chambers, S. P., Caron, P. R., Murcko, M. A. & Wilson, K. P. (1996). *Cell*, **85**, 921–930.
 Townley, R. & Shapiro, L. (2007). *Science*, **315**, 1726–1729.