Crystallization of *Clonorchis sinensis* 26 kDa glutathione S-transferase and its fusion proteins with peptides of different lengths

Young-Hyun Han,[a] Yong-Hak Chung,[a] Tae-Yun Kim,[b] Sung-Jong Hong,[b] Jung-Do Choi[a] and Yong Je Chung[a]*

[a]School of Biological Sciences, Chungbuk National University, Cheongju 360-763, South Korea, and [b]Department of Parasitology, College of Medicine, Chung-Ang University, Seoul 156-756, South Korea

Correspondence e-mail:
chungyj@cbucc.chungbuk.ac.kr

A *Clonorchis sinensis* 26 kDa glutathione S-transferase (*Cs*GST) and its fusion proteins containing 14 and 48 amino-acid peptides at the N-terminus have been crystallized using polyethylene glycol monomethylether 550 as a precipitant. Crystals of the three proteins show very similar crystal properties: they diffract to at least 2.3 Å resolution and belong to the orthorhombic space group $P2_12_12_1$. The unit-cell parameters of *Cs*GST crystals were $a = 66.64\ (1)$, $b = 68.91\ (1)$, $c = 123.41\ (2)$ Å, which are very close to those of the crystals of the two fusion proteins. In addition, *Cs*GST fusion proteins containing varying extents of N-terminal-extended peptides are incorporated into a crystal, indicating that the extended peptides have little effect on crystal packing. These results suggest that the crystallization system of *Cs*GST/peptide fusion protein may be generally applicable to obtain crystals of small peptides.

1. Introduction

Glutathione S-transferases (GSTs) are a group of enzymes that catalyze the conjugation of glutathione (GSH) to a wide variety of electrophilic substrates, resulting in elimination of the GSH adducts from the cell. Such detoxification properties of the enzymes are believed to be responsible to the development of resistance of cells towards drugs, herbicides and pesticides (Hayes & Pulford, 1995). GSTs are also related to many cellular functions, such as their polymorphisms on cancer susceptibility (Carmichael *et al.*, 1988). A number of three-dimensional structures of GSTs have been determined by X-ray crystallography and have revealed that the monomer of all functional GSTs is composed of two domains: a smaller N-terminal α/β-domain and a larger C-terminal α-domain. Although GSTs share this common folding pattern, each structure has characteristic features, particularly concerning the substrate-binding site. In helminths, the enzymes are known as potential chemo- and immuno-therapeutic targets (Ferru *et al.*, 1997; Sexton *et al.*, 1990). So far, crystal structures have been determined for isozymes from *Schistosoma japonicum* (*Sj*GST: Lim *et al.*, 1994; McTigue *et al.*, 1995) and *Fasciola hepatica* (Rossjohn *et al.*, 1997).

One of the major difficulties in determining the structure of proteins by X-ray diffraction is to grow large single crystals: the crystallization procedure still remains an empirical and time-consuming step. Fortunately, recent studies concerning the use of GST-fusion proteins in crystallization (Kuge *et al.*, 1997; Lim *et al.*, 1994; Tang *et al.*, 1999; Ware *et al.*, 1999; Zhang

et al., 1998) provide a clue to a rational approach to solving the above problem. In particular, if the protein segments fused to GST have little effect on crystal packing and the fusion proteins are crystallized under the same conditions used for GST (Tang *et al.*, 1999; Zhang *et al.*, 1998), crystallization is no longer a limiting factor in X-ray structure determination.

We have crystallized a helminth GST, a 26 kDa isozyme from *C. sinensis* (*Cs*GST), in order to determine its three-dimensional structure. During the course of this study we have observed that two *Cs*GST fusion proteins containing 14 and 48 amino-acid peptides at the N-terminus, 14NP*Cs*GST and 48NP*Cs*GST, can be crystallized under the conditions similar to those used for *Cs*GST. In this report, we present a possible application of *Cs*GST fusion proteins as a useful system for growing crystals of small peptides and for determining their structures, together with the results of the crystallization experiments.

2. Materials and methods

2.1. Purification

The 48NP*Cs*GST construct (Hong *et al.*, unpublished work) is a recombinant pBluescript SK(+) derived from a clone purified from λ ZAP library by *in vivo* excision. In this construct, a gene for the N-terminus of *Escherichia coli* β-galactosidase is in frame to the *Cs*GST and is expressed as a fusion peptide. A recombinant plasmid, 14NP*Cs*GST, was constructed by subcloning 48NP*Cs*GST cDNA restricted with *Eco*RI and *Xho*I endonucleases

crystallization papers

into a pMet expression vector. The pMet vector was constructed by re-ligating pET-23b plasmid DNA after *Nde*I/*Bam*HI double digestion and flushing both ends. The coding region of *Cs*GST cDNA was amplified by the polymerase chain reaction (PCR) employing a primer encoding the N-terminus of the enzyme and an antisense primer T7. The *Cs*GST expression plasmid was constructed by subcloning the PCR product double-restricted with *Nde*I and *Xho*I endonucleases into the pET-23b expression vector (Novagen Co.).

Prokaryotic expression host cells, *E. coli* BL21(DE3)pLysS, were transformed with the recombinant plasmid DNAs and grown in LB medium containing ampicillin (50 µg ml⁻¹) at 310 K. Protein expression was induced by adding isopropyl-β-D-thiogalactopyranoside to the culture medium to a final concentration of 1 mM. After an additional 5 h culture, the cells were harvested and disrupted by ultrasonication. The supernatant of the *E. coli* lysate was loaded onto a GSH affinity column. The resin was prepared according to the procedure of Simons & vander Jagt (1977) using epoxy-activated Sepharose 6B (Pharmacia LKB Biotechnology, Uppsala, Sweden). Bound protein was eluted from the column with 15 mM GSH solution. Active fractions were identified by an enzymatic assay (Habig *et al.*, 1974). Homogeneous fractions selected by analyzing SDS–PAGE gels were pooled.

2.2. Crystallization

The sample was dialyzed against 5 mM potassium phosphate buffer pH 7.0 and then concentrated using a Centricon concentrator (Amicon Corporation) until the protein

concentration reached 10–20 mg ml⁻¹. Initial crystallization conditions were screened according to the sparse-matrix method (Jancarik & Kim, 1991) using the hanging-drop vapour-diffusion technique.

As described later, 48NP*Cs*GST is progressively cleaved to smaller proteins. To elucidate the effect of the cleavage on crystallization, the protein was allowed to be cleaved by incubating samples at room temperature for times varying between 12 h and 3 d and the resulting samples (VNP*Cs*GSTs) were used for crystallization experiments. N-terminal sequence analysis of 48NP*Cs*GST incubated for 3 d was carried out using an Applied Biosystems Procise 491 Automatic Sequencer (Perkin–Elmer, Applied Biosystems Division). In order to demonstrate that protein in the crystal retains its intact molecular size, crystals were carefully washed with reservoir solution and dissolved in a phosphate buffer. Only one crystal from each sample was selected for the experiment. The dissolved crystals were then analyzed by SDS–PAGE.

2.3. Preliminary X-ray diffraction analysis

Diffraction data were collected at room temperature on a DIP-2000 imaging-plate detector (MacScience Co.) using Cu $K\alpha$ radiation from a MacScience M06XHF rotating-anode X-ray generator operated at 50 kV and 90 mA. A total of 90 data frames were measured, with an exposure of 15 min and an oscillation angle of 1.0° per frame. The crystal-to-detector distance was set to 120 mm. The raw data were processed to 2.3 Å using *DENZO* and scaled using *SCALEPACK* (Otwinowski & Minor, 1996).

3. Results and discussion

The extended peptide of 48NP*Cs*GST consists of 37 amino-acid residues of the β-galactosidase α-subunit, Met-Thr-Met-Ile-Thr-Phe-Ser-Ala-Gln-Leu-Thr-Leu-Thr-Lys-Gly-Asn-Lys-Ser-Trp-Ser-Ser-Thr-Ala-Val-Ala-Ala-Ala-Leu-Glu-Leu-Val-Asp-Pro-Pro-Gly-Cys-Arg, and 11 amino-acid residues encoded by the 5′-untranslated region of *Cs*GST, Asn-Ser-Ala-Arg-Ala-Ala-Ile-Arg-Glu-Pro-Glu. 14NP*Cs*GST contains the above 11 amino-acid peptide and an extra-peptide, Met-Asp-Pro, at the N-terminus. 48NP*Cs*GST was progressively cleaved to smaller proteins and eventually to a 26 kDa protein in 3 d. It was not possible to derive any sequence data from N-terminal sequence analysis for the resulting 26 kDa

protein, probably because of the N-terminal amino group being blocked by an unidentified chemical group. This suggests that the extended peptide is removed from the fusion protein. Mixtures of *Cs*GST fusion proteins with varying extents of N-terminal extended peptides, VNP*Cs*GSTs, were prepared according to the procedure described in §2.

Large single crystals of *Cs*GST were obtained reproducibly from a 10% PEG monomethylether 550 solution containing 5 mM zinc sulfate in 0.1 M 2-(N-morpholino)ethanesulfonic acid buffer pH 6.5. Rod-shaped crystals appeared within 1 d and grew to dimensions of 0.8 × 0.5 × 0.4 mm in about 3 d. Under these conditions, 14NP*Cs*GST, 48NP*Cs*GST and VNP*Cs*GSTs were crystallized using a similar procedure as in the case of *Cs*GST.

SDS–PAGE analysis of the dissolved crystals of *Cs*GST, 14NP*Cs*GST and 48NP*Cs*GST showed strong bands at the predicted positions on the gel, indicating that crystals contain intact proteins. Interestingly, the dissolved crystals of 48NP*Cs*GST and VNP*Cs*GSTs showed multiple bands (lanes *C*, *D*, *E*, *F* and *G* in Fig. 1), indicating that each crystal contains a mixture of different-sized fusion proteins and partitioning has not occurred during the crystallization.

X-ray diffraction data for the *Cs*GST crystal were collected with 98.6% completeness to 2.3 Å resolution at room temperature. R_{merge} values were 5.3% overall and 18.6% in the last resolution shell between 2.42 and 2.3 Å. Data were also collected for crystals of both 14NP*Cs*GST and 48NP*Cs*GST. The data sets revealed that the crystals belong to the orthorhombic space group $P2_12_12_1$. The unit-cell parameters of *Cs*GST crystals were a = 66.64 (1), b = 68.91 (1), c = 123.41 (2) Å, which are very close to those of 14NP*Cs*GST and 48NP*Cs*GST, with unit-cell parameters a = 66.52 (1), b = 69.02 (2), c = 123.60 (2) Å and a = 66.48 (1), b = 68.98 (2), c = 123.72 (2) Å, respectively. Crystals of VNP*Cs*GST did not show significant changes in unit-cell parameters (data not shown). Assuming one dimer molecule per crystallographic asymmetric unit, a molecular mass of 25.5 kDa and a partial specific volume of 0.74 cm³ g⁻¹, the crystal volume per protein mass (V_M) and the solvent content of *Cs*GST crystals were calculated to be 2.78 Å³ Da⁻¹ and 56%, respectively (Matthews, 1968). The solvent content of 48NP*Cs*GST crystals, with the largest V_M (2.36 Å³ Da⁻¹) of the three proteins being investigated, was estimated to be 48%, which is still typical for protein crystals.
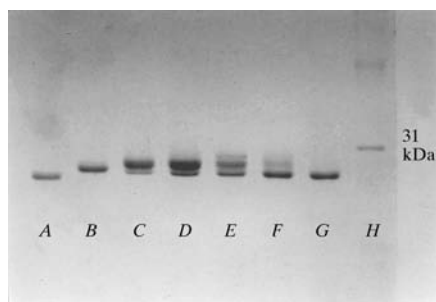
Figure 1
SDS–PAGE analysis of dissolved crystals. Lane *A*, *Cs*GST; lane *B*, 14NP*Cs*GST; lane *C*, 48NP*Cs*GST; lane *D*, 48NP*Cs*GST incubated for 12 h; lane *E*, 48NP*Cs*GST incubated for 24 h; lane *F*, 48NP*Cs*GST incubated for 36 h; lane *G*, 48NP*Cs*GST incubated for 3 d; lane *H*, marker protein (carbonic anhydrase). Samples loaded onto lanes *D*, *E*, *F* and *G* are referred to as VNP*Cs*GSTs. SDS–PAGE was performed on a 16 × 18 × 1.5 cm gel containing 12% acrylamide.

The preparation of proteins using *Sj*GST as a tag protein have advantages in cloning, expressing and purifying the target proteins through glutathione-based affinity chromatography. Prior to crystallization, the proteins fused to *Sj*GST are usually cleaved from *Sj*GST with an appropriate protease and purified *via* additional steps. In contrast to relatively large proteins, small peptides fused to *Sj*GST can be crystallized without further processing (Kim *et al.*, 1999; Kuge *et al.*, 1997; Lim *et al.*, 1994; Tang *et al.*, 1999; Ware *et al.*, 1999; Zhang *et al.*, 1998). The use of fusion proteins may have advantages in obtaining crystals of the peptides fused to *Sj*GST and in determining their structures by employing the molecular-replacement technique (Kuge *et al.*, 1997). In our case, two *Cs*GST-fusion proteins and *Cs*GST are crystallized in the same crystal form under similar conditions. Moreover, VNP*Cs*GSTs are incorporated into a crystal that allows any relative amounts of constituent proteins. These results reflect the fact that *Cs*GST crystals have spaces large enough to accommodate the fused peptides without significant changes in the crystal lattice.

Taken together, it is believed that the crystallization system of *Cs*GST-fusion proteins, like *Sj*GST-fusion proteins, will provide a useful method for obtaining crystals of small peptides and for determining their structures. The structure analysis of *Cs*GST using the molecular-replacement method is currently in progress.

References

Carmichael, J., Forrester, L. M., Lewis, A. D., Hayes, J. D., Hayes, P. C. & Wolf, C. R. (1988). *Carcinogenesis*, **9**, 1617–1621.

Ferru, I., Geroges, B., Bossus, M., Estaquier, J., Delacre, M., Harn, D. A., Tartar, A., Capron, A., Grassmasse, H. & Auriault, C. (1997). *Parasite Immunol.* **19**, 1–11.

Habig, W. H., Pabst, M. J. & Jacoby, W. B. (1974). *J. Biol. Chem.* **249**, 7130–7139.

Hayes, J. D. & Pulford, D. J. (1995). *Crit. Rev. Biochem. Mol. Biol.* **30**, 445–600.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

Kim, S., Shin, D. W., Kim, D. H. & Eom, S. H. (1999). *Acta Cryst.* D**55**, 1601–1603.

Kuge, M., Fujii, Y., Shimizu, T., Hirose, F., Matsukage, A. & Hakoshima, T. (1997). *Protein Sci.* **6**, 1783–1786.

Lim, K., Ho, J. X., Keeling, K., Gilliland, G. L., Ji, X., Ruker, F. & Carter, D. C. (1994). *Protein Sci.* **3**, 2233–2244.

McTigue, M. A., Williams, D. R. & Tainer, J. A. (1995). *J. Mol. Biol.* **246**, 21–27.

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

Otwinowski, Z. & Minor, W. (1996). *Methods Enzymol.* **276**, 307–326.

Rossjohn, J., Feil, S. C., Wilce, M. C. J., Sexton, J. L., Spithill, T. W. & Parker, M. W. (1997). *J. Mol. Biol.* **273**, 857–872.

Sexton, J. L., Milner, A. R., Panaccio, M., Waddington, J., Wijffels, G., Chandler, D., Thompson, C., Wilson, L., Spithill, T. W., Mitchell, G. F. & Campbell, N. G. (1990). *J. Immunol.* **145**, 3905–3910.

Simons, P. C. & vander Jagt, D. L. (1977). *Anal. Biochem.* **82**, 334.

Tang, L., Guo, B., Javed, A., Choi, J. Y., Hiebert, S., Lian, J. B., van Wijnen, A. J., Stein, J. L., Stein, G. S. & Zhou, G. W. (1999). *J. Biol. Chem.* **274**, 33580–33586.

Ware, S., Donahue, J. P., Hawiger, J. & Anderson, W. F. (1999). *Protein Sci.* **8**, 2663–2671.

Zhang, Z., Devarajan, P., Dorfman, A. L. & Morrow, J. S. (1998). *J. Biol. Chem.* **273**, 18681–18684.