Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization and preliminary X-ray analysis of *Escherichia coli* K12 YgjK protein, a member of glycosyl hydrolase family 63

Takashi Tonozuka, Akiko Uechi, Masahiro Mizuno, Kazuhiro Ichikawa, Atsushi Nishikawa and Yoshiyuki Sakano*

Department of Applied Biological Science, Tokyo University of Agriculture and Technology, 3-5-8 Saiwai-cho, Fuchu, Tokyo 183-8509, Japan

Correspondence e-mail: sakano@cc.tuat.ac.jp

Received 13 February 2004
Accepted 21 April 2004

Processing α-glucosidase I, which is classified into glycosyl hydrolase (GH) family 63, hydrolyzes an oligosaccharide precursor of eukaryotic N-linked glycoproteins. Recently, many bacteria have been reported to possess genes for proteins that are homologous to the GH family 63 glucosidases. In this paper, *Escherichia coli* K12 YgjK protein, a member of GH family 63, was overexpressed, purified and crystallized using the vapour-diffusion method. Diffraction data were collected to 1.8 Å resolution and the crystal was found to belong to the monoclinic space group $P2_1$, with unit-cell parameters $a = 88.5$, $b = 137.1$, $c = 60.9$ Å, $\beta = 98.1^\circ$. The V_M value was determined to be 2.1 Å^3 Da^{-1}, which corresponds to the presence of two protein molecules in the asymmetric unit.

1. Introduction

Processing α-glucosidase I (EC 3.2.1.106) specifically hydrolyzes the terminal α1,2-glucosidic linkage of $Glc_3Man_9GlcNAc_2$, which has been identified as an oligosaccharide precursor of N-linked glycoproteins in eukaryotes (Dhanawansa *et al.*, 2002; Dairaku & Spiro, 1997; Moremen *et al.*, 1994). The enzyme catalyzes the first step in trimming the N-linked oligosaccharide precursor and is followed by the action of processing α-glucosidase II, which removes the two α1,3-linked glucose residues. Subsequent modification, which is catalyzed by various glycosidases and glycosyltransferases, produces a wide variety of N-linked oligosaccharides attached to proteins. These reactions have been reported to be responsible for the folding and quality control of newly formed glycoproteins (see reviews by Herscovics, 1999; Wang & Hebert, 2003) and processing α-glucosidase I is a key enzyme for the regulation of N-linked oligosaccharide processing in the eukaryotic cell.

The classification of Coutinho & Henrissat (1999) is widely used for glycosyl hydrolases (GHs; http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html). In this classification system, processing α-glucosidase I falls into GH family 63 and has no homology to any other glycosidase of known three-dimensional structure (Dhanawansa *et al.*, 2002). No three-dimensional structures of this enzyme have yet been reported, probably because the construction of heterologous expression systems for eukaryotic proteins is generally more difficult than for bacterial proteins. Interestingly, many bacteria have been reported to possess genes for proteins that are homologous to the GH family 63 glucosidases, although their physiological roles are not known. Here, we report the crystallization and preliminary X-ray analysis of *Escherichia coli* YgjK protein, the C-terminal part of which exhibits 21% identity to that of processing α-glucosidase I from *Saccharomyces cerevisiae*.

2. Materials and methods

2.1. Construction of the expression vector

The gene manipulations were carried out based on those of Sambrook *et al.* (1989). The YgjK gene was obtained by polymerase chain reaction (PCR) using the genomic DNA of *E. coli* K12 and the oligonucleotides 5′-AAT ATT TCT ACT GGT AAT AAG GAC TGT AAT-3′ and 5′-AAG GCG TTT ACG CCG CAT CCG CCA GCA GTT-3′, which encode the sequences upstream and downstream of the YgjK gene, respectively (Blattner *et al.*, 1997). The amplified fragment was inserted into the pGEM-T vector (Promega) and the 2.4 kbp *Sph*I–*Sac*I fragment was further subcloned into pUC119, resulting in the plasmid pYgjK-119. Construction of an expression plasmid was carried out by oligonucleotide-directed mutagenesis using the QuikChange mutagenesis kit (Stratagene). The plasmid pYgjK-119 was modified with the oligonucleotide 5′-TTT CAC ACA GGA AAC AGC TAT GAA AAT AAA AAC TAT TT-3′ and the same sequence on the opposite strand in order to connect the N-terminal methionine codon for YgjK with the initiation codon of the *lac*Z gene derived from pUC119, resulting in the plasmid pYgjK-SIG.

2.2. Production and purification

Cloned YgjK was prepared from *E. coli* BL21(DE3) cells harbouring the plasmid

pYgjK-SIG. The cells were grown at 303 K in 600 ml Luria–Bertani (LB) medium containing ampicillin (50 µg ml^{-1}) to A_{600} = 0.6–0.8 and then induced with IPTG at a final concentration of 0.1 m*M*. Incubation was then continued overnight. The cells were harvested by centrifugation at 10 000*g* for 5 min, resuspended in 15 ml 20 m*M* Tris–HCl pH 7.5 and disrupted by sonication. The supernatant obtained by centrifugation at 10 000*g* for 15 min was applied onto a HiPrep 16/10 Phenyl FF (high sub) column (1.6 × 10 cm, Amersham) equilibrated with 20 m*M* Tris–HCl pH 7.5 containing 1.9 *M* ammonium sulfate. The protein was eluted using a linear gradient of 1.9–0 *M* ammonium sulfate at a flow rate of 3 ml min^{-1}. The active fractions were collected and dialyzed against 20 m*M* Tris–HCl pH 7.5 and the protein was applied onto a Q-Sepharose 16/10 HP anion-exchange column (1.6 × 10 cm, Amersham) equilibrated with the same buffer. The enzyme was eluted with a linear gradient of 0–0.2 *M* sodium chloride in the same buffer at a flow rate of 3 ml min^{-1}. The purified protein was detected as a single band on SDS–PAGE. The N-terminal amino-acid sequence was determined using a protein sequencer (Applied Biosystems, Model 476A).

2.3. Crystallization and data collection

Crystals were grown by the hanging-drop vapour-diffusion method (McPherson, 1982) at 293 K. The protein solution was prepared at a concentration of 27 mg ml^{-1} in 10 m*M* Tris–HCl pH 7.5. The crystallization drop consisted of 1.5 µl protein solution and an equal volume of crystallization reservoir solution containing 30%(*w*/*v*) PEG 8000, 0.6 *M* magnesium chloride in 100 m*M* Tris–HCl pH 7.0. To perform data collection at cryogenic temperature, the crystal was transferred to the reservoir solution, which was used as the cryoprotectant solution in this study, and then immediately flash-frozen in a stream of nitrogen gas at 100 K.

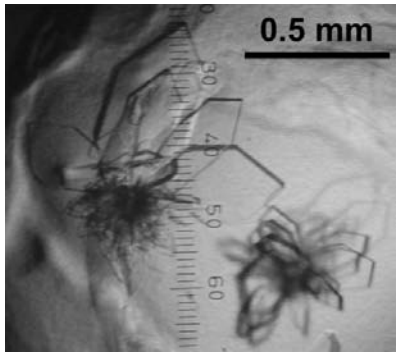

Figure 1
Crystals of *E. coli* YgjK protein.

Diffraction data were collected at the PF-AR NW-12 beamline (Tsukuba, Japan) and the data set was processed with the program *DPS* (Rossmann & van Beek, 1999).

3. Results and discussion

The YgjK protein was overexpressed in *E. coli* harbouring pYgjK-SIG and purified and the N-terminal amino-acid sequence was analyzed. While the deduced N-terminal sequence coded in the original genomic DNA of *E. coli* K12 is MKIKTILTPVTC-ALLISFSAHAANADNY-, the sequence of the purified protein was determined to be NADNY-. This result indicated that the initial N-terminal 23 amino-acid residues, MKIKTILTPVTCALLISFSAHAA, had been proteolytically removed in the *E. coli* BL21(DE3) cells; the molecular weight of the purified protein was calculated to be 85 900 Da. A homology search of the removed N-terminal peptide was implemented in the DDBJ database using the *BLAST* program and found to be homologous to the N-terminal sequences of several protein precursors, such as PilA, a structural subunit of pili, from *E. coli* strain LS3-1 (DDBJ/EMBL/GenBank No. AY082806; 52% identity), and a periplasmic sorbitol-binding protein from *Mesorhizobium loti* (AP003005, 36%). Standard signal peptides of secretory proteins from Gram-negative bacteria have three parts, namely region N (2–15 residues), which is rich in Lys or Arg, region H (>8 residues), which is composed of hydrophobic residues, and region C, which contains Ala or Gly or Ser (Pugsley, 1993). The removed N-terminal peptide also seems to consist of three parts, the Lys-rich MKIKT, the hydrophobic ILTPVTCALLISF and the Ala-rich SAHAA. Therefore, it is likely that the peptide is a signal sequence for secretion and proper folding.

Thin plate-like crystals of YgjK protein reaching dimensions of 0.3 × 0.2 × 0.02 mm grew within a week (Fig. 1). The preliminary crystallographic data obtained under cryo-conditions correspond to a monoclinic crystal and a full set of intensity data to 1.8 Å resolution has been collected. Data-collection statistics and crystal data are summarized in Table 1. A solvent content of 42% (V_M = 2.1 Å^3 Da^{-1}) was calculated using the program *MATTHEWS_COEFF* from the *CCP*4 software suite (Collaborative Computational Project, Number 4, 1994). This indicates that the crystals are expected to contain two molecules per asymmetric unit. We are searching for heavy-atom derivatives for use in the determination of the structure.

Table 1
Data-collection statistics.

Values in parentheses are for the highest resolution shell (1.89–1.80 Å).

Wavelength (Å)	1.0
Temperature (K)	100
Resolution range (Å)	29.5–1.8
No. observed reflections	930373
No. unique reflections	131055
Completeness (%)	98.9 (97.8)
$R_{\rm merge}$†	0.088 (0.207)
$I/\sigma(I)$	5.8 (3.4)
Space group	$P2_1$
Unit-cell parameters	
a (Å)	88.5
b (Å)	137.1
c (Å)	60.9
β (°)	98.1
Unit-cell volume (Å^3)	731062

† $R_{\rm merge} = \sum\sum|I_i - \langle I\rangle|/\sum\langle I\rangle$.

We thank Dr Shigehiro Kamitori for his useful advice and discussions. This study was supported in part by the National Project on Protein Structural and Functional Analyses from the Ministry of Education, Culture, Sports, Science and Technology of Japan and the Novozymes Japan Research Fund. This research was performed with the approval of the Photon Factory Advisory Committee, National Laboratory for High Energy Physics, Tsukuba (2003G110).

References

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. & Shao, Y. (1997). *Science*, **277**, 1453–1474.

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

Coutinho, P. M. & Henrissat, B. (1999). *Recent Advances in Carbohydrate Bioengineering*, edited by H. J. Gilbert, G. Davies, B. Henrissat & B. Svensson, pp. 3–12. Cambridge: Royal Society of Chemistry.

Dairaku, K. & Spiro, R. G. (1997). *Glycobiology*, **7**, 579–586.

Dhanawansa, R., Faridmoayer, A., van der Merwe, G., Li, Y. X. & Scaman, C. H. (2002). *Glycobiology*, **12**, 229–234.

Herscovics, A. (1999). *Biochim. Biophys. Acta*, **1473**, 96–107.

McPherson, A. (1982). *Preparation and Analysis of Protein Crystals*, pp. 82–159. New York: John Wiley.

Moremen, K. W., Trimble, R. B. & Herscovics, A (1994). *Glycobiology*, **4**, 113–125.

Pugsley, A. P. (1993). *Microbiol. Rev.* **57**, 50–108.

Rossmann, M. G. & van Beek, C. G. (1999). *Acta Cryst.* D**55**, 1631–1640.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press.

Wang, T. & Hebert, N. (2003). *Nature Struct. Biol.* **10**, 319–321.