Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Cloning, overexpression, crystallization and preliminary X-ray analysis of a family 1 β-glucosidase from *Streptomyces*

Alicia Guasch,[a]* Miquel Vallmitjana,[b] Rosa Pérez,[a] Enrique Querol,[b] Josep A. Pérez-Pons[b] and Miquel Coll[a]

[a]Institut de Biologia Molecular CSIC Jordi Girona 18-26, 08034 Barcelona, Spain, and [b]Departament de Bioquímica i Biologia Molecular and Institut de Biologia Fonamental, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain

Correspondence e-mail: agmcri@cid.csic.es

Received 30 September 1998
Accepted 22 October 1998

An intracellular β-glucosidase (Bgl3) from *Streptomyces* sp. has been cloned and overexpressed in *Escherichia coli*. The introduction of a His tag at the N-terminal end of the protein has allowed its purification to homogeneity by a single chromatographic step, with yields of 150–200 mg of pure protein per litre of *E. coli* culture. The enzyme (52.6 kDa) is a retaining glycosidase able to hydrolyze a wide range of disaccharides and oligosaccharides and to perform transglycosylation. Crystals of recombinant Bgl3 have been grown from an ammonium sulfate solution using the hanging-drop vapour-diffusion method at 293 K. The crystals belong to the orthorhombic space group *I*222 with unit-cell dimensions a = 101.6, b = 113.4 and c = 187.5 Å at room temperature and contain two molecules per asymmetric unit. A full 1.69 Å resolution diffraction data set (97.7% completeness) has been collected from frozen crystals in a solution containing 30% sucrose, using synchrotron radiation.

1. Introduction

β-Glucosidases (β-glucoside glucohydrolase, E.C. 3.2.1.21) are enzymes which act on carbohydrates and catalyze the transfer of glucosyl groups between oxygen nucleophiles by means of a general acid mechanism with retention of the anomeric configuration and through the formation of a covalent glycosyl–enzyme intermediate (McCarter & Withers, 1994; White & Rose, 1997). Under physiological conditions, such a transfer reaction generally results in the hydrolysis of a β-glucosidic bond from disaccharides, oligosaccharides or conjugated glucosides. Thus, β-glucosidases are widely distributed in nature, playing different roles. Their involvement in enzymatic systems is relevant to plant biomass degradation and molecular pathologies such as Gaucher's disease in humans. Another important feature of β-glucosidases is their transglycosylation activity (*i.e.* transfer of glycosidic residues), which could provide a valuable tool for the specific enzymatic synthesis of sugars and sugar derivatives. On the basis of sequence similarity and hydrophobic cluster analysis, β-glucosidases have been classified in families 1 and 3 of the large group of glycosyl hydrolases, which includes more than 60 families (Henrissat & Davies, 1997). According to this classification, family 1 includes 6-phospho-β-galactosidases (E.C. 3.2.1.85), 6-phospho-β-glucosidases (E.C. 3.2.1.86), lactase-phlorizin hydrolases (E.C. 3.2.1.108) and myrosinases (E.C. 3.2.3.1) in addition to β-glucosidases. With regard to the overall folding, family 1 enzymes are included into the clan or superfamily GH-A, characterized by an $(\alpha/\beta)_8$ barrel structure, whereas no crystallographic data are available for any enzymes belonging to family 3. The GH-A clan contains at least 11 glycosyl hydrolase families representing a wide variety of enzymatic activities which include, among others, endoglucanases, lichenases, glucosylceramidases and xylosidases (Bairoch, 1998). Such a picture means that the α/β barrel fold is a tertiary structure able to accommodate a diversity of both amino-acid sequences and substrate specificities. Hence, work is still necessary to fully understand the elements governing structure–function relationships in this superfamily of glycosyl hydrolases. In recent years, five crystal structures of enzymes belonging to family 1 glycosyl hydrolases have been reported: a cyanogenic β-glucosidase from *Trifolium repens* (Barrett *et al.*, 1995), a phospho-β-galactosidase from *Lactococcus lactis* (Wiesmann *et al.*, 1995), a myrosinase from *Sinapis alba* (Burmeister *et al.*, 1997), a thermostable β-galactosidase from *Sulfolobus solfataricus* (Aguilar *et al.*, 1997) and a β-glucosidase from *Bacillus polymyxa* (Sanz-Aparicio *et al.*, 1998).

The gene encoding Bgl3 glucosidase from *Streptomyces* sp. QM-B814 was initially cloned by functional complementation of a β-glucosidase-negative mutant of *Streptomyces lividans* and the enzyme, as expressed in this host system, was characterized biochemically. It was shown to be a retaining glycosidase with an exo-like action pattern, releasing glucose units from the non-reducing end of diverse disaccharides and oligosaccharides (Pérez-Pons *et al.*, 1994). Upon sequence-similarity analysis, it

has been classified into the family 1 glycosyl hydrolases, showing identity indexes from 35 to 55% with respect to different members of that family. Protein crystallization trials were initially carried out with protein purified from crude cell extracts of recombinant *S. lividans* by six chromatographic steps with an overall yield of 10% (unpublished results). Large-scale fermenter cultures were necessary to obtain amounts of pure protein sufficient for extensive crystallization screening. Crystals were obtained from several conditions by the vapour-diffusion method but all of them exhibited very poor X-ray diffraction properties, probably owing to their small size. In order to overcome the time-consuming and low-yielding protein preparation process from *Streptomyces*, attempts to express Bgl3 in *E. coli* were undertaken, including extracellular expression by fusion to several signal peptides from *E. coli* and *Bacillus licheniformis*.

In this communication, we report the crystallization and preliminary X-ray analysis of the Bgl3 glucosidase of *Streptomyces* sp. QM-B814 as expressed in *E. coli* using the pET system (Novagen), which allowed very high yields of enzyme after a one-step purification scheme. This work constitutes the first step towards a detailed three-dimensional structure analysis of this enzyme at high resolution, which will afford new insights into the overall architecture of the GH-A clan and the structural elements determining the fine tuning of substrate specificity in the large glycosyl hydrolase family 1.

2. Experimental

2.1. Cloning, expression and purification of Bgl3 from *E. coli*

The β-glucosidase gene from *Streptomyces* sp. QM-B814, including its downstream region which may function as a transcription terminator since it contains two inverted repeats (CGGTGCGGCAC and GGCCCGCCCCCG starting at 8 and 37 nucleotides from the stop codon, respectively), was cloned into the T7 polymerase expression vector pET-21d(+) (Novagen) and overexpressed in *E. coli* strain BL21(DE3) (Novagen). Likewise, a His-tag coding sequence was fused to the start codon (GTG) of the Bgl3 reading frame obtaining the recombinant plasmid pET21-HBG3, which yields a protein with an extended N-terminal formed by the sequence MHHHHHHGIH and a deduced molecular mass of 53.6 kDa.

In order to obtain reproducible yields of protein, each production batch was started by transforming competent *E. coli* BL21(DE3) cells with the plasmid pET21-HBG3. One transformant colony grown in LB solid medium containing ampicillin (100 µg ml^{-1}), was picked, inoculated into 3 ml of 2xYT broth containing ampicillin (200 µg ml^{-1}) (Sambrook *et al.*, 1989) and shaken overnight at 310 K to an OD_{595} of about 0.6; aliquots of this culture were used to inoculate larger volumes of the same growth medium (usually 200 ml) that were shaken (300 rev min^{-1}) for 8–10 h at 310 K. Following such a procedure, expression of recombinant Bgl3 was achieved constitutively without addition of isopropyl-β-thiogalactopyranoside (IPTG). The cells were harvested and disrupted essentially according to the protocol described in the pET System Manual (Novagen); the sonication step was omitted (see §3). For protein purification, cell-free extracts were centrifuged (39000*g*, 30 min at 277 K) and the supernatant (soluble fraction) applied to a 5 ml Hi-Trap Chelating Sepharose (Pharmacia) column previously charged with Ni^{2+}. Chromatography was carried out according to the manufacturer's specifications and Bgl3-containing fractions were pooled, concentrated and desalted by ultrafiltration at 277 K using 50 m*M* sodium phosphate (pH 7.0) as desalting buffer. Using this procedure a homogeneous (>99%) protein sample was obtained, as judged by sodium dodecyl sulfate–polyacrylamide-gel electrophoresis (SDS–PAGE) and Coomassie Brilliant Blue R-250 staining. The protein concentration was determined by the dye-binding method of Bradford (1976) using BSA as a standard. The pure enzyme was kept at 277 K for short periods or stored at 253 K in the presence of 45% glycerol for prolonged periods.

2.2. Crystallization

Initial crystallization conditions were obtained by the hanging-drop vapour-diffusion method (McPherson, 1982) with a broad screening according to the conditions described by Jancarik & Kim (1991) and Cudney *et al.* (1994) included in Crystal Screening Kits I and II (Hampton Research, California, USA). The droplets were prepared by mixing 2 µl of protein solution (12 mg ml^{-1}) with an equal volume of mother liquor. The trials were carried out at a constant temperature of either 277 or 293 K. Another set of trials was prepared using a trypsinized Bgl3 sample (see §3). Time-course trypsin (Boehringer) digestion was performed at a protease:Bgl3 ratio of 1:100 in 45 m*M* Tris–HCl buffer (pH 8.1) at 310 K. Reactions were stopped by adding 1 m*M* Pefabloc SC (Boehringer) and analyzed by means of SDS–PAGE and Coomassie Blue staining. The trypsin cleavage site was determined by microsequencing the first ten N-terminal amino-acid residues from each digestion product using the Edman degradation method combined with mass-spectrometry analysis (MALDI–TOF system).

2.3. X-ray crystallography

Diffraction data were collected both at 293 K using an in-house rotating-anode X-ray source with $\lambda = 1.5418$ Å and at 103 K on the beamline BW7A of the EMBL outstation at the Deutches Elektronensynchrotron (Hamburg, Germany), $\lambda = 0.99$ Å. In both cases a MAR Research 300 mm imaging plate was used as detector. Determination of unit-cell parameters, space group, integration of reflection intensities and data scaling were performed using *DENZO* and *SCALEPACK* (Otwinowski, 1993).

3. Results and discussion

The crystallization of the Bgl3 glucosidase has proved to be more difficult than expected, probably owing to intrinsic structural features of the protein and not to the presence of impurities or sample heterogeneity. Nevertheless, it deserves to be mentioned that the inclusion of a sonication step during the production of the crude extract seems to interfere negatively with the crystal growth, probably as a consequence of the solubilization of misfolded Bgl3 protein leading to sample microheterogeneity. In any case, the cloning and overexpression of Bgl3 in *E. coli* using the pET vector and the fusion of a His tag at the N-terminus allowed improvement in both the availability of large amounts of protein and the simplicity of its purification. In fact, pure protein yields increased from about 70 mg per 15 l fermenter culture of *S. lividans* (and six chromatographic steps) to 150–200 mg l^{-1} of *E. coli* culture following the one-step purification procedure described in §2. In contrast, the fusion of the His tag to the C-terminal end of Bgl3 glucosidase was deleterious for its expression, as was any attempt to obtain secretion by fusing different signal peptides to the N-terminal end.

As stated in §1, at the beginning of this work we obtained small poorly diffracting

crystals of native Bgl3 purified from *S. lividans*. Attempts were made to improve the crystal size and diffraction quality by seeding, changes in buffer, the use of additives such as alcohols, organic solvents and detergents, and the presence of substrates or inhibitors. None of these trials led to better and useful results. Likewise, many attempts to obtain crystals of *E. coli*-expressed native Bgl3 were carried out, but all were unsuccessful. It is known that in some cases limited proteolysis improves both crystal growth and stability (Chitarra *et al.*, 1995). Thus, protease digestions were checked by using different enzymes, obtaining the best and reproducible results with trypsin reactions. After 1 h digestion two bands of about 35 and 20 kDa were observed, and no further proteolysis was detected upon extending the reaction for up to 6 h. Following the methodology described in §2, we determined that the major trypsin-cleavage sites correspond to Ser315 and Arg330, yielding two fragments of 33560 and 16608 Da consistent with the mass estimated by SDS–PAGE analysis. Furthermore, trypsinized samples were loaded onto a Sephacryl S-100 gel-filtration column previously equilibrated with 0.2 *M* ammonium acetate buffer (pH 6.0) and only one peak eluted from the column. This peak was active using *p*-nitrophenyl-β-D-glucopyranoside as a substrate, indicating that trypsin digestion gives a stable and functional Bgl3 core enzyme. Upon sequence alignment between Bgl3 and family 1 glycosidases whose structures are already known (see §1), it can be inferred that the trypsin-cleavage sites are situated within a region forming a loop. In addition, such a loop would be larger in the *Streptomyces* enzyme.

Table 1
Data collection and statistics.

Temperature (K)	293	103
Unit-cell dimensions (Å)	a = 101.6, b = 113.4, c = 187.5	a = 96.7, b = 111.3, c = 185.3
Resolution range (Å)	40–2.7	40–1.69
Number of unique reflections	26565	108818
Completeness (%)		
All data	88.2	97.7
Highest resolution shell†	84.3	68.4
Average redundancy	2.3	4.4
$\langle I/\sigma(I)\rangle$		
All data	23.4	34.3
Highest resolution shell†	22.7	5.87
$R_{\rm merge}$‡ (%)		
All data	9.7	3.8
Highest resolution shell†	11.3	23.2

† Highest resolution shell was 2.8–2.7 and 1.72–1.69 Å for 293 and 103 K crystals, respectively. ‡ $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$.

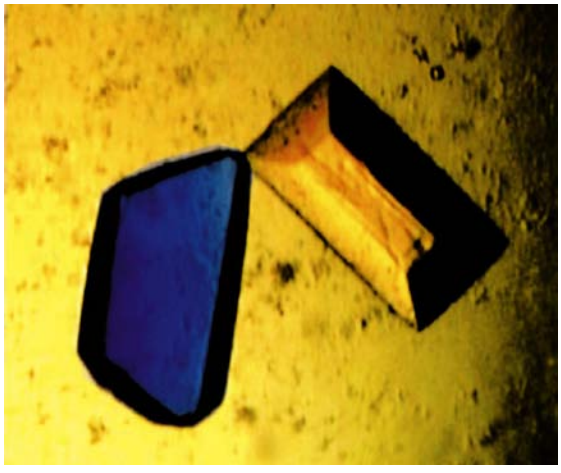

Figure 1
Crystals of trypsinized β-glucosidase, grown in 1.9 *M* ammonium sulfate, 0.1 *M* HEPES (pH 7.5) at 293 K; dimensions are approximately 0.4 × 0.4 × 0.2 mm.

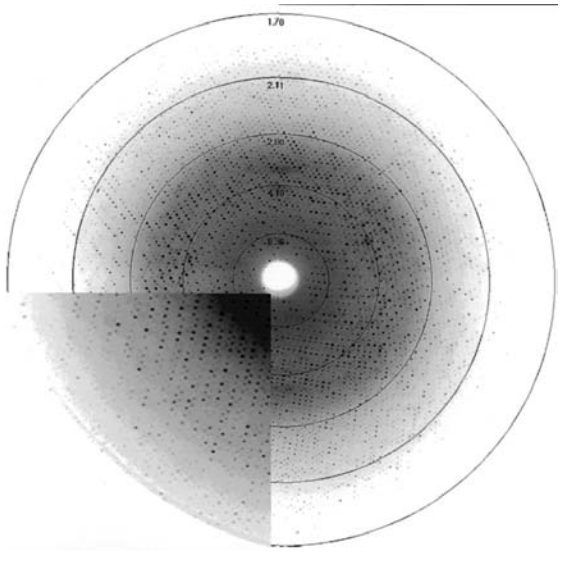

Figure 2
1° rotation image obtained from a frozen crystal using synchrotron radiation. Resolution rings are labelled in Å.

Crystals were obtained in 1.9 *M* ammonium sulfate, 0.1 *M* HEPES (pH 7.5) at 293 K. Single well shaped colourless bipyramidal crystals appeared in about three weeks (Fig. 1), reaching a full size of about 0.4 × 0.4 × 0.2 mm after four months. The crystals diffracted beyond 2.7 Å resolution using a rotating-anode source at 293 K. A data set was collected, but a significant change of the cell dimensions (5 Å in a) was observed after 50 frames. The indexing procedure carried out with *DENZO* (Otwinowski, 1993) indicated that the crystals belong to space group *I*222, with unit-cell dimensions of a = 101.6, b = 113.4 and c = 187.5 Å. Flash-freezing after crystal transfer to a solution containing 30% sucrose and 2.25 *M* ammonium sulfate (pH 7.5) improved crystal stability and diffraction quality (Fig. 2). Using synchrotron radiation, frozen crystals diffracted beyond 1.69 Å resolution and the unit-cell dimensions were a = 96.7, b = 111.3 Å and c = 185.3 Å. Table 1 shows the data-collection statistics. The unit-cell volume is consistent with two monomers in the asymmetric unit, yielding an acceptable specific volume (V_m) of 2.48 Å^3 Da^{-1} and a corresponding solvent content of 40% (Matthews, 1968). The structure determination by molecular replacement using *Trifolium repens* cyanogenic β-glucosidase (PDB entry 1CBG; Barrett *et al.*, 1995) is in progress.

The *E. coli* expression system has also proven to be useful for the overproduction of Bgl3 point mutants obtained by site-directed mutagenesis, and several different mutants have been produced. The crystallization and structure determination of these mutant enzyme forms will provide useful tools in the evaluation and understanding of the principles that determine substrate specificity. With this aim, we have also prepared a series of co-crystals with non-hydrolyzable substrate analogues and heavy-atom containing inhibitors.

This work was supported by the Ministerio de Educación y Ciencia (grants PB95-0224 to MC and BIO 97-0511-CO2-01 to EQ) and the Generalitat de Catalunya [Centre de Referència en Biotecnologia

(CERBA) and grant 1997SGR-275 to MC]. Data collection at EMBL-DESY was supported by the EU Large Installations Project CHGECT-93-0040. We thank F. Canals for his assistance in performing mass spectrometry and amino-acid sequencing analyses. MV is the recipient of a predoctoral scholarship from CERBA.

References

Aguilar, C. F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M. & Pearl, L. H. (1997). *J. Mol. Biol.* **271**, 789–802.

Bairoch, A. (1998). *SWISS-PROT Protein Sequence Data Bank*, http://expasy.hcuge.ch/cgi-bin/lists?glycosid.txt.

Barrett, T., Suresh, C. G., Tolley, S. P., Dodson, E. J. & Hughes, M. A. (1995). *Structure*, **3**, 951–960.

Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.

Burmeister, W. P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S. & Henrissat, B. (1997). *Structure*, **5**, 663–675.

Chitarra, V., Souchon, H., Spinelli, S., Juy, M., Beguin, P. & Alzari, P. M. (1995). *J. Mol. Biol.* **248**, 225–232.

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. (1994). *Acta Cryst.* D**50**, 414–423.

Henrissat, B. & Davies, G. (1997). *Curr. Opin. Struct. Biol.* **7**, 637–644.

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

McCarter, J. D. & Withers, S. G. (1994). *Curr. Opin. Struct. Biol.* **4**, 885–892.

McPherson, A. (1982). *The Preparation and Analysis of Protein Crystals*. New York: John Wiley.

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

Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.

Pérez-Pons, J. A., Cayetano, A., Rebordosa, X., Lloberas, J., Guasch, A. & Querol, E. (1994). *Eur. J. Biochem.* **223**, 557–565.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*. New York: Cold Spring Harbor Laboratory Press.

Sanz-Aparicio, J., Hermoso, J. A., Martínez-Ripoll, M., Lequerica, J. L. & Polaina, J. (1998). *J. Mol. Biol.* **275**, 491–502.

White, A. & Rose, D. R. (1997). *Curr. Opin. Struct. Biol.* **7**, 645–651.

Wiesmann, C., Beste, G., Hengstenberg, W. & Schulz, G. E. (1995). *Structure*, **3**, 961–968.