Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Characterization of different crystal forms of the α-glucosidase MalA from *Sulfolobus solfataricus*

Heidi Asschenfeldt Ernst,[a] Martin Willemoës,[a,b] Leila Lo Leggio,[a] Gordon Leonard,[c] Paul Blum[d] and Sine Larsen[a,c]*

[a]Centre for Crystallographic Studies, Department of Chemistry, University of Copenhagen, Denmark, [b]Department of Biological Chemistry, University of Copenhagen, Denmark, [c]European Synchrotron Radiation Facility (ESRF), Grenoble, France, and [d]Beadle Centre for Genetics, University of Nebraska, Nebraska, USA

Correspondence e-mail: sine.larsen@esrf.fr, sine@ccs.ki.ku.dk

Received 14 September 2005
Accepted 27 October 2005
Online 5 November 2005

MalA is an α-glucosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus*. It belongs to glycoside hydrolase family 31, which includes several medically interesting α-glucosidases. MalA and its selenomethionine derivative have been overproduced in *Escherichia coli* and crystallized in four different crystal forms. Microseeding was essential for the formation of good-quality crystals of forms 2 and 4. For three of the crystal forms (2, 3 and 4) full data sets could be collected. The most suitable crystals for structure determination are the monoclinic form 4 crystals, belonging to space group $P2_1$, from which data sets extending to 2.5 Å resolution have been collected. Self-rotation functions calculated for this form and for the orthorhombic ($P2_12_12_1$) form 2 indicate the presence of six molecules in the asymmetric unit related by 32 symmetry.

1. Introduction

α-Glucosidases (EC 3.2.1.20; α-D-glucoside glucohydrolases) are exo-acting enzymes which cleave the terminal glucose moiety from the non-reducing end of their substrates. This group of enzymes play essential roles in carbohydrate metabolism and in glycoprotein processing. They catalyze the release of glucose from oligosaccharides and storage polymers such as starch and glycogen. Regulation of α-glucosidase activity helps to control glucose levels and therefore several of these enzymes are among drug targets for type 2 diabetes (Cheng & Fantus, 2005). Deficiency or malfunction of the human acid α-glucosidase causes accumulation of glycogen in the lysosomes, a condition known as glycogen-storage disease II (Pompe's disease). The endoplasmic reticulum α-glucosidases catalyze processing of N-linked oligosaccharides in the quality-control system for glycoprotein folding and maturation; inhibitors of these enzymes have potential as antiviral agents (Mehta *et al.*, 1998).

In the glycoside hydrolase classification (Henrissat, 1991; Henrissat & Bairoch, 1993; http://afmb.cnrs-mrs.fr/cazy/CAZY/index), α-glucosidases are found mainly in two families: GH13 and GH31. Most α-glucosidases from higher organisms belong to the GH31 family. This family comprises proteins displaying a range of different enzymatic activities (α-glucosidase, α-xylosidase, α-1,4-glucan lyase, 6-α-glucosyltransferase and 3-α-isomaltosyltransferase) with very broad substrate specificities (Frandsen & Svensson, 1998). Structural information for the GH31 family is currently limited to an α-xylosidase from *Escherichia coli* (Lovering *et al.*, 2005).

MalA is an α-glucosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus*, with a substrate preference for maltose and maltooligosaccharides (Rolfsmeier & Blum, 1995). It consists of 693 amino-acid residues and has a molecular weight of 80.4 kDa. Since proteins from thermophilic organisms are often more amenable to crystallization, MalA was chosen as a crystallization target representing the α-glucosidase specificity in GH31.

2. Materials and methods

2.1. Construction of a plasmid for overproduction of MalA

Standard molecular-biology methods for cloning and manipulation of DNA were used (Sambrook *et al.*, 1989). The entire coding region of malA was obtained by PCR with Turbo *Pfu* DNA polymerase

Table 1
Crystal data and data-collection statistics.

Values in parentheses refer to the highest resolution shell.

	Form 2	Form 3	Form 4	Form 4b
	MalA	SeMet-MalA	SeMet-MalA	MalA
Crystallization conditions	4 mg ml^{-1} MalA, 11% PEG 4000, 0.1 *M* sodium acetate pH 4.1, 0.2 *M* sodium citrate, microseeding (form 2)	3 mg ml^{-1} SeMet-MalA, 31% PEE797†, 0.1 *M* sodium acetate pH 4.0, 10 m*M* DTT‡	4 mg ml^{-1} SeMet-MalA, 11% PEG 4000, 0.1 *M* sodium acetate pH 4.1, 0.2 *M* sodium citrate, microseeding (form 2)	4 mg ml^{-1} MalA, 10% PEG 4000, 0.1 *M* sodium acetate pH 4.2, 0.2 *M* sodium citrate, 0.5% BOG§, microseeding (form 2)
Cryocooling conditions	25% glycerol	—	25% glycerol, annealed	25% glycerol, annealed
Beamline	ID14-2, ESRF	ID14-2, ESRF	ID23-1, ESRF	ID14-1, ESRF
Detector	ADSC Q4R CCD	ADSC Q4R CCD	MAR Mosaic 225 CCD	ADSC Q4 CCD
Wavelength (Å)	0.9330	0.9330	0.9791	0.9340
Space group	$P2_12_12_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters				
a (Å)	97.3	321.5	103.2	100.1
b (Å)	192.8	158.2	173.6	174.4
c (Å)	281.7	320.6	154.1	144.0
β (°)		119.9	108.0	109.1
Molecules per ASU	4–8	21–48	4–9	4–8
Resolution range (Å)	35–3.0 (3.1–3.0)	35–4.0 (4.2–4.0)	35–2.55 (2.65–2.55)	35–2.50 (2.63–2.50)
Unique reflections	105696	231876	167413	154886
Observed reflections	787671	842228	1262798	398868
$R_{\rm mrgd\text{-}F}$¶ (%)	5.6 (16.1)	14.7 (23.5)	6.3 (18.2)	9.6 (30.5)
$\langle I/\sigma(I)\rangle$	20.3 (7.6)	7.2 (4.9)	16.8 (7.1)	11.7 (5.3)
Completeness (%)	98.6 (95.4)	97.4 (95.6)	99.2 (98.5)	95.2 (95.8)
Redundancy	7.4 (7.1)	3.6 (3.7)	7.5 (7.5)	2.6 (2.3)

† PEE797, pentaerythritol ethoxylate (15/4 EO/OH). ‡ DTT, dithiothreitol. § BOG, β-octylglucopyranoside. ¶ $R_{\rm mrgd\text{-}F}$ reflects the quality of the reduced amplitudes considering the gain in accuracy resulting from data redundancy (Diederichs & Karplus, 1997).

(Stratagene) used according to the supplier's recommendations with the vector pB122 (Rolfsmeier *et al.*, 1998) as a template. The oligonucleotide primers SSMALA-5, 5′-CG*GGATCC*TAAAAATATACGAGAACAAAGGC-3′, and SSMALA-3, 5′-AA*CTGCAG*CTACTCTAGGTTAATCTTTCCCC-3′, were used. These introduce unique *Bam*HI and *Pst*I restriction sites (shown in italics) at the 5′- and 3′-ends of the PCR product, respectively. Following digestion with the restriction enzymes *Bam*HI and *Pst*I, the PCR product was ligated into the vector pUHE 23-2 (Lutz & Bujard, 1997), resulting in the plasmid pMW800. *E. coli* strain NF1830 (Jensen *et al.*, 1992) was transformed with pMW800 to allow IPTG (isopropyl-β-D-thiogalactopyranoside) controlled production of MalA. Owing to the cloning strategy, the first four amino acids of MalA (^{1}MQTI) are changed to ^{1}MRIL in the recombinant protein.

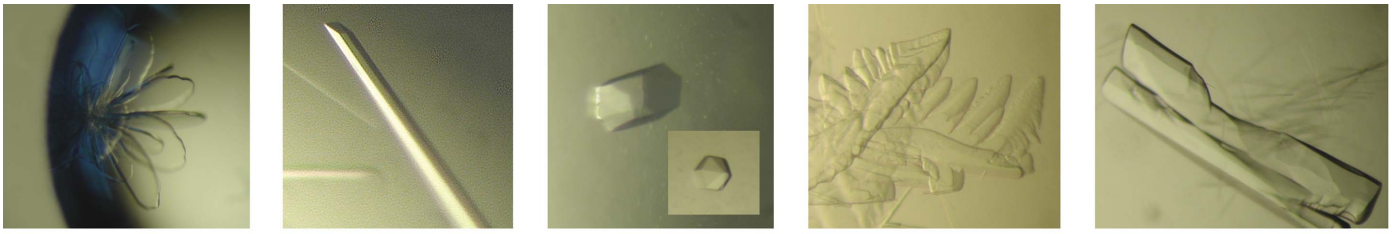
2.2. Overexpression and purification

A 5 ml overnight culture of the MalA-overproducing strain grown at 310 K in LB medium (Sambrook *et al.*, 1989) supplemented with ampicillin (100 µg ml^{-1}) was transferred to 1 l of the same medium and growth was continued at 310 K. When the culture reached an OD_{436} of 0.8–1, IPTG was added to a final concentration of 1 m*M* and incubation of the culture was continued overnight. The cells were harvested by centrifugation in a Sorvall centrifuge (SLA3000 rotor) at 8000 rev min^{-1} for 5 min. The cell paste was washed in 50 m*M* Tris–HCl pH 7.5, 100 m*M* NaCl, recovered by centrifugation as above and resuspended in 40 ml of the same buffer. The cells were lysed by sonication and the cell debris was removed by centrifugation in the Sorvall centrifuge (SS34 rotor, 14 000 rev min^{-1}, 30 min). The cell extract, diluted to 80 ml in the same buffer, was stirred gently while placed in a water bath preheated to 358 K. When the temperature of the cell extract reached 348 K it was allowed to stand for 10 min and then put on ice. The precipitate was removed by centrifugation. The supernatant was then brought to 50% saturation with ammonium sulfate on ice and the precipitate was recovered by centrifugation. The precipitate was redissolved in 4 ml 20 m*M* Tris–HCl pH 7.5, 20 m*M* NaCl and dialysed against the same buffer at 277 K. The dialysed protein was applied onto a 5 ml HiTrap Q column (Pharmacia) and eluted with a gradient of 20–500 m*M* NaCl in 20 m*M* Tris–HCl pH 7.5. MalA eluted as a single peak around 50 m*M* NaCl. Subsequently, the sample was loaded onto a HiLoad 26/60 Superdex 200 gel-filtration column (Pharmacia) equilibrated with 50 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl and eluted using a flow rate of 0.5 ml min^{-1}. The above procedure yielded about 10 mg of protein which was more than 95% pure as judged from SDS–PAGE. For preparation of the selenomethionine-substituted MalA (SeMet-MalA), the overproducing strain was grown as previously described (Duyne *et al.*, 1993; Doublié, 1997) and the enzyme was purified as described above.

The substitutions at the N-terminus of the recombinant protein were confirmed by N-terminal sequencing (^{1}MRILKIYENK). Peptide mapping of a tryptic digest of MalA by MALDI–TOF mass spectrometry resulted in about one-third sequence coverage, although the terminal peptides were not seen. MalA and SeMet-MalA were assayed for α-glucosidase activity as described by Rolfsmeier & Blum (1995) using *p*-nitrophenyl-α-D-glucopyranoside as substrate.

2.3. Crystallization

Crystals were grown at room temperature by the hanging-drop vapour-diffusion method using protein stock solutions of 2–5 mg ml^{-1} in 20 m*M* Tris–HCl pH 7.5, 20 m*M* NaCl. Drops consisting of 2 µl protein and 3 µl reservoir solution were equilibrated over 1000 µl reservoirs. In the initial solubility (Stura *et al.*, 1992) and sparse-matrix (Hampton Research Crystal Screen 1; Jancarik & Kim, 1991) screens, organic polymers generally produced crystalline precipitates. A variety of morphologically different crystals were subsequently obtained in a tailor-made screen with lower concentration of polyethylene glycol 4000 [PEG 4000, 10%(*w*/*v*)], 0.2 *M* salt



(*a*) (*b*) (*c*) (*d*) (*e*)

Figure 1
Crystals of MalA. (*a*) Form 1, maximum dimensions 0.1 × 0.2 × 0.02 mm; (*b*) form 2, maximum dimensions 0.3 × 0.05 × 0.05 mm; (*c*) form 3, maximum dimensions 0.15 × 0.1 × 0.1 mm; (*d*) form 4, irregular formations with maximum thickness around 0.04 mm; (*e*) form 4b, maximum dimensions 0.6 × 0.1 × 0.05 mm. Crystallization conditions for crystal forms 2–4b are given in Table 1. The form 1 crystals shown in (*a*) were grown from 0.1 *M* Tris–HCl pH 8, 0.2 *M* ammonium sulfate, 10% PEG 4000 using a protein concentration of 3.9 mg ml^{-1}.

(ammonium sulfate, sodium dihydrogen phosphate or sodium citrate) and covering the pH interval 4–8. Rosettes of roundish plates (form 1, Fig. 1*a*) were obtained at high pH (7–8.5) with either ammonium sulfate or sodium dihydrogen phosphate. At low pH (4–5) and in the presence of citrate, bundles of thin needles were formed (form 2). Owing to the apparent preference for polymeric precipitating agents, separate screens based on different types of organic polymers, pentaerythritol propoxylates and pentaerythritol ethoxylates (Gulick *et al.*, 2002), were designed. Crystals of hexagonal shape (form 3, Fig. 1*c*) grew from high concentrations of pentaerythritol ethoxylate (15/4 EO/OH; PEE797) at pH 4. Each crystal form was optimized using fine grid screens with variations in pH, concentration of the salt, polymer or protein and in the relative volume of protein solution and reservoir in the drop. Usually, crystals formed within 3–10 d. Microseeding proved essential for reproducible growth of single crystals of form 2 (Fig. 1*b*). In a few cases, crystals displaying different morphologies (Figs. 1*d* and 1*e*) formed prior to the form 2 crystals in the optimization experiments where microseeding was used. Optimized crystallization conditions for crystals of forms 2–4b that could be used for data collection are summarized in Table 1.

2.4. Data collection and processing

Initial diffraction tests were performed at I711, MAXLAB (Lund, Sweden). Diffraction data were collected at cryogenic temperatures using solutions consisting of reservoir composition supplemented with 25% glycerol for cryoprotection. One exception is the form 3 crystals, which were flash-cooled directly from the mother liquor, as PEE797 itself acts as a cryoprotectant at high concentrations. Various beamlines (ID14-1, ID14-2, ID14-4 and ID23-1) at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) were used for further testing and data collection (Table 1). Data were processed with *XDS* (Kabsch, 1993) and scaled using *XSCALE*. Statistics for the best data set collected for each crystal form are shown in Table 1.

3. Results and discussion

Construction of a new overexpression system in *E. coli* allowed the production of recombinant MalA in large quantities. The specific activities of MalA and SeMet MalA were comparable to previous reports on the native and recombinant enzyme (Rolfsmeier & Blum, 1995; Rolfsmeier *et al.*, 1998). A variety of crystal forms (Fig. 1) have been obtained and characterized by X-ray diffraction (Table 1). Only minor differences were observed in the crystallization behaviour of SeMet MalA and unsubstituted MalA and the diffraction quality of the resulting crystals was similar. Space groups were assigned based on averaging statistics and systematic absences, and the number of molecules in the asymmetric unit was estimated assuming a solvent content of between 30 and 70%. The form 1 crystals were obtained under a range of different conditions, mostly at pH values above 7 where the protein carries a negative net charge (pI $\simeq$ 5.5). Diffraction test images showed unit-cell parameters very similar to those of the orthorhombic form 2 crystals (Table 1), but severe disorder in at least one direction hampered data collection. In contrast, the form 2 crystals displayed ordered diffraction and complete data sets extending to approximately 3 Å resolution could be collected. The form 3 crystals diffracted X-rays beyond 2.5 Å, but a very large unit cell (Table 1) caused considerable problems with overlapping reflections and data could only be processed to around 4 Å. Despite the pseudo-hexagonal geometry of the unit cell (a = 321.5, b = 158.2, c = 320.6 Å, β = 119.9°), the diffraction pattern only obeys monoclinic symmetry ($P2_1$). This corresponds to 21–48 molecules per asymmetric unit, which together with the problems in data acquisition renders this crystal form unsuitable for structure determination. Notably, the crystals of irregular shape (Figs. 1*d* and 1*e*) obtained in a few instances after microseeding with the form 2 crystals diffracted to a resolution limit beyond 2.5 Å. These crystals both belong to the monoclinic space group $P2_1$ and show similar, but not identical, unit-cell parameters (Table 1). Consequently, they have been denoted form 4 (SeMet MalA; Fig. 1*d*) and form 4b (MalA; Fig. 1*e*), respectively. Crystal-packing densities suggest that the asymmetric unit contains four to eight molecules. The self-rotation functions for the form 2 and the form 4 data, calculated with the programs *MOLREP* (Vagin & Teplyakov, 1997), *AMoRe* (Navaza, 1994) and *POLARRFN* (Collaborative Computational Project, Number 4, 1994), revealed the presence of non-crystallographic twofold and threefold symmetry axes. In both cases, a group of NCS symmetry elements compatible with 32 point-group symmetry could be identified, which indicates the presence of six molecules in the asymmetric unit. Data collected from the SeMet-substituted protein (form 4) did not contain sufficient anomalous signal for phasing. This could be a consequence of incomplete substitution. Structure determination is therefore being pursued by molecular replacement using the recently reported structure of the GH31 α-xylosidase YicI from *E. coli* (PDB code 1xsi; Lovering *et al.*, 2005) as a search model.

The authors would like to thank Dorthe Boelskifte for assistance in protein purification, Christine Finnie and Birte Svensson (Carlsberg Laboratory, Valby, Denmark) for help in peptide mapping by mass spectrometry and Flemming Hansen for help in crystal handling. Provision of beamtime at MAXLAB (Lund, Sweeden) and the European Synchrotron Radiation Facility (ESRF, Grenoble, France) is gratefully acknowledged. This study was supported by the European Union under the Cell Factory of the Fifth Framework Program, QLK3-2001-02400 (NEPSA), the Danish National

Research Foundation and the Danish Natural Sciences Research Council.

References

Cheng, A. Y. Y. & Fantus, I. G. (2005). *Can. Med. Assoc. J.* **172**, 213–226.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
Doublié, S. (1997). *Methods Enzymol.* **276**, 523–529.
Frandsen, T. P. & Svensson, B. (1998). *Plant Mol. Biol.* **37**, 1–13.
Gulick, A. R., Horswill, J. B. & Thoden, J. C. (2002). *Acta Cryst.* D**58**, 306–309.
Henrissat, B. (1991). *Biochem. J.* **280**, 309–316.
Henrissat, B. & Bairoch, A. (1993). *Biochem. J.* **293**, 781–788.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Jensen, K. F., Andersen, J. T. & Poulsen, P. (1992). *J. Biol. Chem.* **267**, 17147–17152.
Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
Lovering, A. L., Lee, S. S., Kim, Y.-W., Withers, S. & Strynadka, N. C. J. (2005). *J. Biol. Chem.* **280**, 2105–2115.
Lutz, R. & Bujard, H. (1997). *Nucleic Acids Res.* **25**, 1203–1210.
Mehta, A., Zitzmann, N., Rudd, P. M., Block, T. M. & Dwek, R. A. (1998). *FEBS Lett.* **430**, 17–22.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Rolfsmeier, M. & Blum, P. (1995). *J. Bacteriol.* **177**, 482–485.
Rolfsmeier, M., Haseltine, C., Bini, E., Clark, A. & Blum, P. (1998). *J. Bacteriol.* **180**, 1287–1295.
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.
Stura, E. A., Nemerow, G. R. & Wilson, I. A. (1992). *J. Cryst. Growth*, **122**, 273–285.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.