Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of 20K endoglucanase from *Melanocarpus albomyces* at 1.8 Å resolution

Jarkko Valjakka* and Juha Rouvinen

Department of Chemistry, University of Joensuu, PO Box 111, 80101 Joensuu, Finland

Correspondence e-mail: jarkko.valjakka@joensuu.fi

Received 13 September 2002
Accepted 23 January 2003

PDB Reference: 20K endoglucanase, 1l8f, r1l8fsf.

The crystal structure of the 20K endoglucanase from the thermophilic fungus *Melanocarpus albomyces* (Ma20k) has been determined. The structure was refined to 1.8 Å resolution using data obtained at 120 K. Ma20k belongs to glycoside hydrolase family 45. The three-dimensional structures of endoglucanase V (EGV) from the fungus *Humicola insolens* and of an endoglucanase from *H. grisea* var. *thermoidea* have previously been determined. The overall structure of Ma20k consists of a six-stranded β-barrel domain similar to that found previously in family 45 endoglucanases. The flexible loop between strands V and VI, which was disordered in the uncomplexed structures of the *Humicola* endoglucanases but was ordered in complexed structures of EGV, is found to be well ordered in the native structure of Ma20k. The structure of Ma20k allows comparison between thermophilic and mesophilic proteins of family 45 and different principles for thermostability are discussed.

1. Introduction

Cellulases catalyse hydrolysis of the β-(1,4) glycosidic bonds of cellulose, the most abundant polysaccharide on earth. Generally, a variety of bacteria and fungi convert cellulose to soluble sugars as a source of energy for cells. Cellulases can be used in manufacturing, *e.g.* in the wood industry for treatment of fibres to facilitate further processing, drainage properties *etc.* In the textile industry, the biostoning of denims is a commercial application (Vehmaanperä *et al.*, 1997). Cellulases can be divided into two groups, endoglucanases (EC 3.2.1.4) and cellobiohydrolases (EC 3.2.1.91). So far, the structures of only a few endoglucanases from family 45 have been determined. Coordinates are available for the native enzyme from *Humicola insolens* (PDB code 2eng) and its complexes with cellobiose (PDB code 3eng) and with cellohexaose (PDB code 4eng) (Davies *et al.*, 1993, 1995, 1996); coordinates for a native endoglucanase (PDB code 1hd5) from *Humicola grisea* var. *thermoidea* are also available. Here, we report the crystallization, structure determination and structure of a new endoglucanase belonging to the family 45 glycoside hydrolases, the 20K endoglucanase from the thermophilic fungus *Melanocarpus albomyces* (Ma20k). Endoglucanase V cleaves the β-(1,4) glycosidic bonds of cellulose with an inversion mechanism (Davies *et al.*, 1995, 1996); therefore, a similar reaction mechanism can be suggested for Ma20k and also for the endoglucanase from *H. grisea* var. *thermoidea* (structure not published). Ma20k is compared with the previously determined endoglucanase structures and possible reasons for the better thermostability of Ma20k are briefly discussed.

2. Materials and methods

2.1. Crystallization

The enzyme was a kind gift from Röhm Enzyme Finland Oy. Crystals of Ma20k were grown using the hanging-drop vapour-phase diffusion method. Initial drops consisted of 1 µl protein solution (9 mg ml^{-1} in 12.5 m*M* Tris–HCl pH 7.2, 12.5 m*M* EDTA, 0.01% sodium azide) and 1 µl reservoir solution (100 m*M* sodium acetate buffer pH 5.10 and 5% PEG 8000). The drops also contained 1 µl of additive solution (30% dextran sulfate). To avoid degradation of the enzyme, it was essential to maintain the pH at above 4. Crystals appeared at 293 K after two weeks.

2.2. Data collection

Diffraction data were collected on a Rigaku RU200HB rotating-anode generator equipped with Osmic Confocal optics and an R-AXIS II imaging-plate system. The crystal-to-detector distance was 135 mm. The data set was measured with 0.7° oscillation frames and the oscillation time for each frame was 20 min. In order to reduce osmotic shock, crystals were harvested with stepwise transfer *via* a series of solutions containing increasing concentrations of the cryoprotective agent (10, 20 and 30%

Table 1
Data collection and refinement statistics.

Outer shell statistics (1.86–1.80 Å) are given in parentheses.

Data collection	
Resolution (Å)	99.0–1.80
Completeness (%)	95.6 (88.3)
Mosaicity (°)	0.354
R_{merge} (%)	9.1 (25.1)
Unique/total reflections	16780/78044
Average $I/\sigma(I)$	14.2
$I > 3\sigma$ (I) (%)	75.8 (43.3)
Refinement statistics and model quality	
Resolution range (Å)	500–1.80
No. of non-H atoms	1566
No. of residues	207
No. of water molecules	205
R factor/R_{free} (%)	19.48/24.70
Space group	$P4_12_12$
Unit-cell parameters (Å)	a = 47.44, b = 47.44, c = 156.81
R.m.s.d. bond lengths (Å)	0.005
R.m.s.d. bond angles (°)	1.35
Ramachandran plot, residues in	
Most favoured region (%)	87.7
Additional allowed region (%)	12.3
Average B factors (Å^2)	
Protein	14.7
Solvent	30.8
R.m.s.d. in B values (Å^2)	
Main-chain atoms	1.119
Side-chain atoms	2.036

glycerol). Crystals were flash-frozen in a cold stream at 120 K (using an Oxford Cryosystems Cryostream cooler). Data were indexed and processed with *DENZO* and the space group was determined with *XPREP* (*SHELXTL* software package). Data from individual image plates were scaled and reflection intensities were merged using *SCALEPACK* (Otwinowski, 1993). The crystals belonged to space group $P4_12_12$, with unit-cell parameters a = 47.44, c = 156.81 Å and one molecule in the asymmetric unit. Data were collected to 1.8 Å resolution. Table 1 summarizes the data-collection statistics. Crystallization of the same enzyme has recently been reported by Hirvonen & Papageorgiou (2002). The space group was the same ($P4_12_12/P4_32_12$), but the unit cell for this crystal form was clearly larger (unit-cell parameters a = 47.3, c = 177.3 Å). Moreover, the diffraction resolution was lower (2.2 Å) using synchrotron radiation.

2.3. Structure solution and refinement

Molecular replacement was carried out using the *AMoRe* program (Navaza, 1994). The structure of native EGV (PDB code 2eng; Davies *et al.*, 1996) was used as the search model for rotation and translation calculations. The model was used in a rotation search using data in the resolution range 8–3 Å with a sphere radius of 20 Å. The correlation coefficient for the rotation solution was 20.8% (next peak 13.5%) and that for the translation solution was 52.7% (next peak 15.1%). Rigid-body refinement gave an R factor of 53.2%. Data refinement, including individual temperature-factor refinement, was performed using *CNS* (Brünger *et al.*, 1998). Map inspection and manual rebuilding were performed with *O* (Jones *et al.*, 1991). Water molecules were added manually as they became apparent in the map. Water molecules which had a temperature factor of lower than 60 Å^2 were kept in the model. The final structure of Ma20k contains 207 residues and 205 water molecules. Electron-density maps indicated a distinct double conformation for the side chain of residue Asp120. The alternate conformations were included in the final refinement calculations. The final structure of Ma20k has an R factor of 19.48% and an R_{free} of 24.70% for all data to 1.8 Å resolution. As reported by *PROCHECK* (Laskowski *et al.*, 1993), all residues lie in or close to the allowed regions of the Ramachandran plot. Coordinates and structure factors of the structure are available from the PDB (PDB code 1l8f; Berman *et al.*, 2000). A summary of the refinement results is given in Table 1.

3. Results and discussion

A sequence comparison of the endoglucanases from *H. insolens*, *H. grisea* var. *thermoidea* and *M. albomyces* is shown in Fig. 1. It shows an identity of ~75% between Ma20k and the endoglucanases of known structure. The sequence of Ma20k, consisting of 207 amino acids, terminates just

```
                10        20        30        40        50
1l8f ANGQSTRYWD CCKPSCGWRG KGPVNQPVYS CDANFQRIHD FDAVSGCE-G
2eng ADGRSTRYWD CCKPSCGWAK KAPVNQPVFS CNANFQRITD FDAKSGCEPG
1hd5 ADGKSTRYWD CCKPSCGWAK KAPVNQPVFS CNANFQRLTD FDAKSGCEPG

                60        70        80        90        100
1l8f GPAFSCADHS PWAINDNLSY GFAATALSGQ TEESWCCACY ALTFTSGPVA
2eng GVAYSCADQT PWAVNDDFAL GFAATSIAGS NEAGWCCACY ELTFTSGPVA
1hd5 GVAYSCADQT PWAVNDDFAF GFAATSIAGS NEAGWCCACY ELTFTSGPVA

                110       120       130       140       150
1l8f GKTMVVQSTS TGGDLGSNHF DLNIPGGGVG LFDGCTPQFG GLPGARYGGI
2eng GKKMVVQSTS TGGDLGSNHF DLNIPGGGVG IFDGCTPQFG GLPGQRYGGI
1hd5 GKKMVVQSTS TGGDLGSNHF DLNIPGGGVG IFDGCTPQFG GLPGQRYGGI

                160       170       180       190       200
1l8f SSRQECDSFP EPLKPGCQWR FDWFQNADNP SFTFERVQCP EELVARTGCR
2eng SSRNECDRFP DALKPGCYWR FDWFKNADNP SFSFRQVQCP AELVARTGCR
1hd5 SSRNECDRFP DALKPGCYWR FDWFKNADNP SFSFRQVQCP AELVARTGCR

                210
1l8f RHDDGGFAVF KAP
2eng RNDDGNFPAV QIP
1hd5 RNDDGNFPAV
```

Figure 1
The amino-acid sequence of Ma20k (1l8f) from *M. albomyces* and the sequences of endoglucanase V (2eng) from *H. insolens* and the endoglucanase from *H. grisea* var. *thermoidea* (1hd5). Different amino acids are shown in bold and the flexible loop between strands V and VI is underlined.

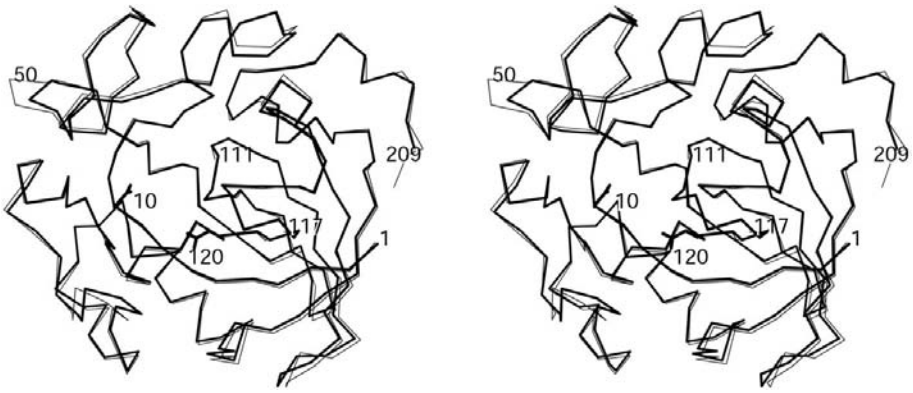

Figure 2
Stereoview of the C$^\alpha$-atom trace of the 20K endoglucanase from *M. albomyces* (thick lines) and of endoglucanase V from *H. insolens* (thin lines). The catalytic residues Asp10 and Asp120 are shown. The largest differences between the two proteins are in loops 47–50 and 111–117. The C$^\alpha$-atom superimposition was performed using *O* (Jones *et al.*, 1991) and the figure was drawn using *SETOR* (Evans, 1993).

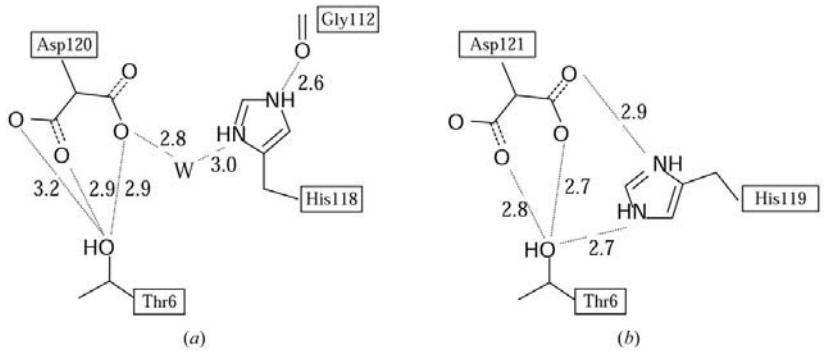

Figure 3
Hydrogen bonding around the catalytic donor aspartate. This residue has two alternate conformations, which are similar to the native structure of endoglucanase V. (*a*) Asp120 of the 20K endoglucanase from *M. albomyces* (Ma20k). (*b*) Asp121 of endoglucanase V from *H. insolens* (EGV). The distances (Å) between non-H atoms are given.

before the linker and the cellulose-binding domain area in EGV. Ma20k has one deletion at position 49. The overall structure of Ma20k consists of a six-stranded β-barrel and its folding is similar to that of the native *Humicola* endoglucanases (PDB codes 2eng and 1hd5). A superimposition of Ma20k and the other endoglucanases is shown in Fig. 2. The root-mean-square differences between the structures are 0.5 Å (calculated from the C^{α}-atom positions).

The active sites of the endoglucanases lie in a large groove on the protein surface. The endoglucanases from *Humicola* (2eng and 1hd5) and *Melanocarpus* have similar catalytic residues, Asp10 (general base) and Asp120 (proton donor), which were found at similar positions in the active site. The catalytic aspartates are on either side of an aromatic Tyr8, which packs against the glucose ring of the substrate during catalysis. The side chain of Asp10 has similar hydrogen bonds to those observed in 2eng and 1hd5. Asp120 of Ma20k has two alternate side-chain conformations; one makes two hydrogen bonds to Thr6 and the other makes one water-mediated hydrogen bond to His118. Furthermore, His118 of Ma20k endoglucanase forms a direct hydrogen bond to Gly112, which is located in the glycine-rich loop between strands V and VI. Fomation of this hydrogen bond is enabled by the different orientation of His118, which is possible because Ala75 on strand III is smaller than the corresponding residue Ser76 in 2eng and 1hd5. In the complexed structures of EGV, the hydroxyl group of Thr6 orientates so as to interact with the catalytic Asp120, which has therefore been suggested to be a catalytic proton donor (Davies *et al.*, 1995). Also in the complexed structures of EGV, the side chain of Thr6 forms a direct hydrogen bond to His119. This interaction does not exist in Ma20k (Fig. 3).

The loop (Gly111-Gly-Asp-Leu-Gly-Ser116) between strands V and VI was disordered in the native EGV structure but is ordered in the complexed EGV structures and the native Ma20k structure. Structure comparison (Fig. 4) shows that the side chain of His118 has a different orientation in Ma20k compared with the other endoglucanase structures, so it is probable that this hydrogen bond with Gly112 has made the structure of the loop more rigid. In Ma20k, the side chain of Gln4 is shorter than Arg4 in 2eng and 1hd5 and therefore is not in close contact with either the side chain of His118 or the flexible loop regions. Gln4 in Ma20k also makes a new hydrogen bond to Asn117 and this new interaction may also stabilize the loop flexibility. A large difference between the loops occurs at the positions of residues Asp113 and Leu114. The side chain of Asp113 makes two internal hydrogen bonds to the main-chain N atoms of Leu114 and Gly115. In addition, Gly115 forms a new hydrogen bond to Ser77; this interaction is absent in 2eng and 1hd5 as they contain Ala78 at this location.

In conclusion, the structures of Ma20k and EGV are highly similar, but the thermostability of Ma20k is significantly higher than that of EGV. It is likely that only exiguous structural modifications have enhanced the thermostability. The amino-acid distribution in the enzymes is similar and there are no significant differences between the residues critical for thermostability (Vogt *et al.*, 1997). The occurences of arginine, serine, threonine and tyrosine are similar in Ma20k and EGV. Structural comparison shows that Ma20k has all seven disulfide bonds in similar positions and only one loop, that between strands I and II, is shorter than in EGV. Hydrogen bonds play an important role in increased stability (Vogt *et al.*, 1997; Kumar *et al.*, 2000). There are some different hydrogen bonds in Ma20k compared with EGV. These may play important roles in the stabilization of the structure. One of them, between the main-chain carbonyl of Asp171 and the side chain of Gln174, is located near the end of the α_b helix (residues 160–171). Another hydrogen bond, between the side chains of Gln79 and Ser83, stabilizes the apex of a loop between strands III and IV. This loop is near the flexible region 111–116. The importance of the number of ion pairs has also been reported by Vogt *et al.* (1997). Ma20k has a few more polar residues than EGV and new ion pairs exist in positions Asp42–Arg19 and Asp203–Arg185. Furthermore, Asp57 makes an ion pair with His58 in Ma20k, instead of Arg32 as in EGV. Residues Glu91 and Arg185 in EGV have been replaced by Ala90 and Glu184 in

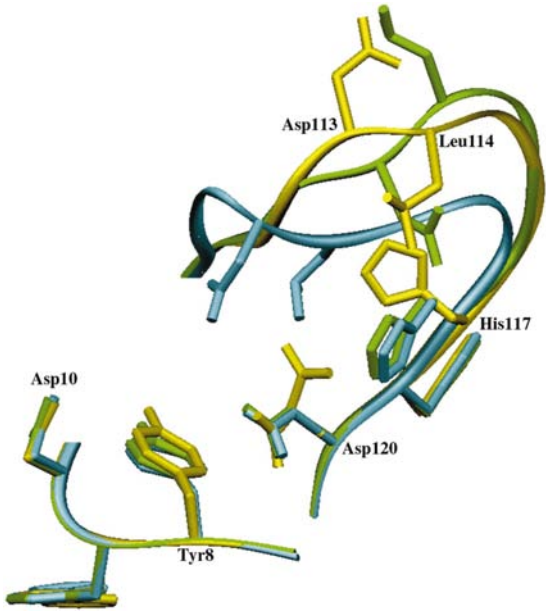

Figure 4
Superimposition of the loops between strands V and VI of the 20K endoglucanase from *M. albomyces* (Ma20k) and the complex structures of endoglucanase V from *H. insolens* (EGV). The native Ma20k is in yellow and EGV complexed with cellobiose and cellohexaose is in green and blue, respectively. The catalytic aspartates Asp10 and Asp120 and structurally important residues are labelled as in Ma20k. The superimposition of the loops was performed using *O* (Jones *et al.*, 1991) and the figure was drawn using *SETOR* (Evans, 1993).

Ma20k; therefore, it does not have the ion pairs in these positions that are present in EGV. In spite of the different ion pairs in Ma20k, the polar surface area has not increased. According to the comparison of Ma20k and EGV, the thermostability difference between mesophilic and thermophilic enzymes in endoclucanase family 45 arises from small modifications in the structures.

References

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Davies, G. J., Dodson, G. G., Hubbard, R. E., Tolley, S. P., Dauter, Z., Wilson, K. S., Hjort, C., Mikkelsen, J. M., Rasmussen, G. & Schulein, M. (1993). *Nature (London)*, **365**, 362–364.

Davies, G. J., Dodson, G. G., Tolley, S. P., Dauter, Z., Wilson, K. S., Rasmussen, G. & Schulein, M. (1996). *Acta Cryst.* D**52**, 7–17.

Davies, G. J., Tolley, S. P., Henrissat, B., Hjort, C. & Schulein, M. (1995). *Biochemistry*, **34**, 16210–16220.

Evans, S. V. (1993). *J. Mol. Graph.* **11**, 134–138.

Hirvonen, M. & Papageorgiou, A. C. (2002). *Acta Cryst.* D**58**, 336–338.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kumar, S., Tsai, C.-J. & Nussinov, R. (2000). *Protein Eng.* **13**, 179–191.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Otwinowski, Z. (1993). *DENZO and SCALEPACK. Data Processing and Scaling Programs*. Yale University, New Haven, Connecticut, USA.

Vehmaanperä, J., Elovainio, M., Haakana, H., Joutsjoki, V., Lantto, R., Mäntylä, A. Paloheimo, M., Suominen, P., Londesborough, J. & Miettinen-Oinonen, A. (1997). Patent No. WO9714804.

Vogt, G., Woell, S. & Argos, P. (1997). *J. Mol. Biol.* **269**, 631–643.