FEBS Letters 554 (2003) 462–466 FEBS 27827

Structural insights into the processivity of endopolygalacturonase I from *Aspergillus niger*

Gertie van Pouderoyen[a], Harm J. Snijder[a], Jacques A.E. Benen[b], Bauke W. Dijkstra[a,*]

[a] *Laboratory of Biophysical Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands*
[b] *Fungal Genomics, Laboratory of Microbiology, Wageningen University, Dreyenlaan 2, 7603 HA Wageningen, The Netherlands*

Received 4 September 2003; revised 10 October 2003; accepted 10 October 2003

First published online 29 October 2003

Edited by Hans Eklund

Abstract Endopolygalacturonase I is a processive enzyme, while the 60% sequence identical endopolygalacturonase II is not. The 1.70 Å resolution crystal structure of endopolygalacturonase I reveals a narrowed substrate binding cleft. In addition, Arg96, a residue in this cleft previously shown to be critical for processivity, interacts with the substrate mimics glycerol and sulfate in several well-defined conformations in the six molecules in the asymmetric unit. From this we conclude that both Arg96 and the narrowed substrate binding cleft contribute to retaining the substrate while it moves through the active site after a cleavage event has occurred.

Key words: Endopolygalacturonase; Processivity; X-ray crystallography; β-Helix; *Aspergillus niger*

1. Introduction

Pectin is a major component of plant cell walls. It has a complex structure, built up of blocks of homogalacturonan, consisting of α(1,4)-linked D-galacturonic acid residues and rhamnogalacturonan, which is a repeat of the dimeric α(1,2)-L-rhamnose-α(1,4)-D-galacturonic acid unit. Polygalacturonases hydrolyze the α(1,4)-glycosidic bonds in the homogalacturonan regions of pectin. *Aspergillus niger* secretes seven highly homologous endopolygalacturonases (endopolygalacturonases I, II and A–E), which randomly attack the polymeric substrate [1–4]. All these polygalacturonases belong to glycosyl hydrolase family 28, a family that also includes rhamnogalacturonases [5,6].

One of the enzymatic characteristics in which the *A. niger* endopolygalacturonases differ is the processive behavior of endopolygalacturonases I, A, C, and D. Processive enzymes do not release the polymer substrate after the hydrolysis reaction, but feed it through the active site cleft for the next cleavage event. Processivity is apparent from product progression analyses on polymeric substrates and from specific bond cleavage frequencies of oligogalacturonates of defined chain length. The processivity of polygalacturonase I was shown to be governed by Arg96 at subsite −5. Changing this residue into a serine yielded a mainly non-processive enzyme [7]. Here we present the 1.7 Å crystal structure of endopolygalacturonase I, the first structure of a processive enzyme from glycosyl hydrolase family 28.

2. Materials and methods

2.1. Protein purification, crystallization, and X-ray data collection

Endopolygalacturonase I was isolated and purified in the same way as described earlier for endopolygalacturonase II [8]. Crystals were grown at room temperature using the hanging drop method. The drops were made by mixing 2 μl of protein solution (8.8 mg/ml in 50 mM sodium citrate buffer, pH 3.0) and 2 μl of reservoir solution (18–20% (w/v) PEG4000, 0.2 M ammonium sulfate, 0–10 mM $NiSO_4$ and 50 mM sodium citrate buffer, pH 3.0). Plates and clusters of plate-like crystals grew in 1–2 weeks. The addition of $NiSO_4$ is not essential but reduces the formation of clustered plates and enhances the formation of thicker plates. The crystals were transferred to a solution of 35% (w/v) PEG4000, 10% (v/v) glycerol, 0.2 M ammonium sulfate and 50 mM sodium citrate buffer, pH 3.0 by a step-wise increase of the PEG4000 and glycerol concentrations. Subsequently the crystals were flash-frozen and stored in liquid nitrogen until data collection. The crystals belong to space group $P1$ with cell dimensions $a = 68.0$ Å, $b = 84.1$ Å, $c = 96.0$ Å and cell angles $\alpha = 114.3°$, $\beta = 98.0°$, $\gamma = 89.8°$. They diffract to 1.70 Å resolution.

Data were collected on beam line ID14-2 at the ESRF synchrotron in Grenoble, with an ADSC Q4 CCD detector using an X-ray wavelength of 0.934 Å (Table 1). Data were processed with DENZO and SCALEPACK [9]. Reduction to structure factor amplitudes was performed with TRUNCATE [10].

2.2. Structure determination

The structure of endopolygalacturonase I was solved by molecular replacement, using a model created by the 3D-PSSM server [11] based on the structure of the 60% sequence identical endopolygalacturonase II (PDB code 1CZF [12]). The molecular replacement program EPMR [13] found six molecules (correlation coefficient 0.493 (30–3 Å)) yielding a solvent content of 43.0%. The six molecules were rigid body refined and simulated annealed with CNS [14]. The resulting electron density map was very clear and manual rebuilding using the programs O [15] and QUANTA (Accelrys, San Diego, CA, USA) was done. The improved model was further refined with REFMAC5 [16]. Water molecules were added. Clear electron density was present for *N*-glycosylation at Asn246, *O*-glycosylation at Ser44 and Ser46, 16 sulfate ions, and five glycerol molecules. The refinement statistics and information about the final refined structure are listed in Table 1. The coordinates have been deposited in the PDB bank under accession number 1NHC.

3. Results and discussion

3.1. Overall structure of endopolygalacturonase I

Endopolygalacturonase I folds into a right-handed parallel β-helical structure comprising 10 complete turns (Fig. 1A), very similar to *A. niger* endopolygalacturonase II [12]. The

*Corresponding author. Fax: (31)-50-3634800.
E-mail address: bauke@chem.rug.nl (B.W. Dijkstra).

0014-5793/03/$22.00
doi:10.1016/S0014-5793(03)01221-3

Table 1
Data collection and refinement statistics

Data collection		
Resolution (Å)		37–1.70
Total observations/unique observations		430 145/205 268
Completeness (last shell (1.74–1.70 Å)) (%)		97.5 (95.6)
R_{merge} (last shell (1.74–1.70 Å)) (%)		4.3 (38.4)
I/σ (last shell (1.74–1.70 Å))		22.2 (2.1)
Refinement statistics		
R-factor/free *R*-factor (%)		17.5/20.9
R.m.s. deviation from ideality:	Bond lengths (Å)	0.015
	Bond angles (°)	1.775

Contents of refined structure

Polypeptide chain	A	B	C	D	E	F	Total
Number of amino acid residues	336	336	336	336	336	336	2016
Number of sugar residues:							
N-Acetylglucosamine	2	2	1	2	1	1	9
Mannose	2	6	2	3	2	1	16
Number of water molecules							1355
Number of sulfate ions	2.5	2	3.5	2	3.5	2.5	16
Number of glycerol molecules	1	1	0	1	1	1	5

enzyme crystallizes with six molecules in the asymmetric unit, called A, B, C, D, E, and F. Except for the two N-terminal residues and some solvent-exposed side chains the molecules are clearly defined in density. Superposition of the six molecules (excluding the two N-terminal residues) results in r.m.s. differences of 0.09 Å (molecules A and F) to 0.28 Å (molecules D and E) with an average r.m.s. difference of 0.18 Å for 334 Cα atoms.

3.2. Glycosylation

In all six molecules in the asymmetric unit electron density was observed extending from Asn246, in which *N*-linked sugar residues could be modeled. In molecule B the most extensive glycosylation (two GlcNAc and four mannoses) is present (Figs. 1A and 2). This glycan chain hydrogen bonds with neighboring endopolygalacturonase I molecules (Glu298 of molecule E with Man1 O4, and Lys358 of molecule D with Man4 O3). Molecules A, C, D, E, and F contain visible density for two GlcNAc, one GlcNAc, two GlcNAc and one mannose, one GlcNAc and one GlcNAc residue, respectively. Furthermore, two *O*-glycosylation sites were found in the structure of endopolygalacturonase I, at Ser44 and Ser46. At both positions the electron density is compatible with a single covalently bound mannose residue. Only in molecule F no sugar density of a bound mannose was visible on Ser44.

3.3. Comparison with family 28 glycosyl hydrolases

The structure of endopolygalacturonase I (336 amino acids, numbered from 33 to 368) is very similar to that of endopolygalacturonase II (335 amino acids, numbered from 28 to 362) [12]. There is a one amino acid insertion at position 126 and the r.m.s. difference of the other 335 corresponding Cα atoms is 0.8 Å. Endopolygalacturonase I is also similar to the other known glycosyl hydrolase family 28 structures [17]. For instance, the r.m.s. difference with the *Aspergillus aculeatus* polygalacturonase [18] is 0.8 Å (for 334 Cα atoms) and 1.4 Å with the endopolygalacturonase from *Stereum purpureum* [19] (for 320 Cα atoms). Comparison with other family 28 structures [20–22] shows higher r.m.s. differences.

3.4. Glycerol and several sulfate ions bind in the substrate binding cleft

In total 16 sulfate ions and five glycerol molecules were observed in the electron density maps at the surface of the six endopolygalacturonase molecules. Two sulfate ion binding sites are present at the same position in all six molecules. The first one is between His229, Arg262 and Lys264. The sulfate ion is bound between the positively charged nitrogen atoms of these residues, with nitrogen–oxygen distances ranging from 2.7 to 4.6 Å. This site is close to the putative +1 substrate binding subsite [3,19] and this sulfate ion might mimic the position of the carboxylate group of the sugar residue bound at this subsite. Sulfate ions have more often been observed to occupy the position of a sugar substrate carboxylate group [23].

The other sulfate ion present in all six molecules is located at the N-terminus of the small helix, where the helix dipole stabilizes its negative charge [24]. It interacts with the backbone amides of Ala40 and Ser41, and the Oγ of Ser39. This binding site is far from the active site and the substrate binding cleft. This is also the case for two sulfate ions present between molecules A and C and between molecules E and F, which mediate crystal contacts between the backbone amides of Tyr87 and the NH1 nitrogen atoms of Arg118 of both neighboring molecules (distances of about 2.9 Å).

In two of the endopolygalacturonase I molecules (C and E) a sulfate ion is bound near Arg96 (Fig. 3C,E; in the other four a glycerol molecule occupies this position (Fig. 3A,B,D,F). In addition, in molecule E a glycerol molecule is present opposite Arg96 (Fig. 3E). These binding sites are near substrate binding subsite −5 and the sulfate ion and/or glycerol may mimic the binding of the (negatively charged) oxygen atoms of the substrate. Glycerol molecules have more often been observed to bind in sugar binding sites (e.g. in xylanase [25], Slt35 [26], myrosinase [27] and β-hexosaminidase [28]).

3.5. The role of Arg96 in processivity

Site-directed mutagenesis experiments have shown that Arg96 is of key importance for the processive behavior of this enzyme [7]. Changing this arginine into a serine removes

the enzyme's processivity. Conversely, when the equivalent serine (91) in endopolygalacturonase II was mutated into an arginine the mutant enzyme became processive. It has been proposed that the positively charged arginine binds the negatively charged substrate better than the serine and that it facilitates the movement of the retained substrate through the active site after cleavage.

The structure fully supports this hypothesis. Arg96 interacts with bound glycerol and/or sulfate ions (see above). Moreover, Arg96 shows several well-defined conformations in the different endopolygalacturonase molecules (Fig. 3A–G). This suggests that Arg96 can bind differently positioned sugar residues preventing release, and at the same time is flexible enough to promote sliding and to guide the retained substrate through the substrate binding cleft towards the active site. This would be in analogy to the *Clostridium cellulolyticum* cellulase Cel48F, a processive enzyme that slides along cellulose polymers, catalyzing multiple rounds of hydrolysis before dissociating from its substrate. This enzyme binds the polymeric substrate in two different sets of binding sites which are shifted by half the length of a glucosyl residue [29,30]. In this way the energy barrier to translocate the substrate is reduced and sliding and processivity is promoted. Substrate binding experiments to substantiate this hypothesis for endopolygalacturonase I have so far been unsuccessful, unfortunately.

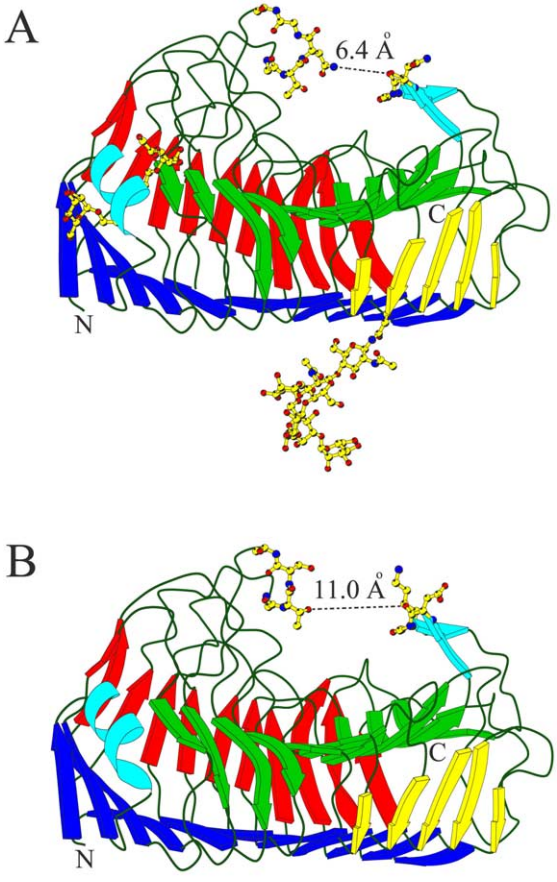

Fig. 1. The three-dimensional structure of (A) endopolygalacturonase I and (B) endopolygalacturonase II with the N- and C-termini indicated, viewed onto β-sheet PB2a (yellow). PB2b (blue) is the bottom β-sheet. PB1 is shown in green and PB3 is shown in red. The active site (Asp186, Asp207 and Asp208) is visible at the top, between the T1 loop regions (on the right side) and the T3 loop regions (on the left side). The glycosylation of endopolygalacturonase I and the loops bordering the active site cleft are shown in ball-and-stick representation (endopolygalacturonase I residues 124–128 (left) and 299–301 (right), endopolygalacturonase II residues 121–123 (left) and 293–295 (right)). The smallest distance at the entry of the active site cleft is indicated with a dashed line and the value is shown [34]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

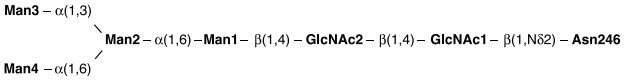

Fig. 2. *N*-Glycosylation at Asn246.

3.6. Narrowed active site cleft

Processive enzymes have been classified according to the degree of enclosure of the substrate [31]. Two broad categories are discerned: enzymes that fully enclose their substrate, and enzymes that only partly cover the substrate binding cleft. While in the former case an obvious explanation can be given for such an enzyme's processivity, this is not the case for enzymes that only partially enclose their substrate. It does appear, however, that processive enzymes tend to enclose their substrates more than non-processive enzymes [31–33].

Endopolygalacturonases I and II form a pair of processive/non-processive enzymes with very similar structures (r.m.s. difference of 0.8 Å, see above). Both have a rather open active site cleft from which substrates could easily diffuse away. Nevertheless, the cleft in endopolygalacturonase I is narrower than that of endopolygalacturonase II (6.4 Å across compared to 11 Å in endopolygalacturonase II, see Fig. 1A,B). This narrowing is caused by the insertion of an asparagine after Thr125 in a loop above the active site. In addition the side chain of Asn126 may play a role in hydrogen bonding the substrate.

4. Conclusions

The structure of endopolygalacturonase I from *A. niger* shows that Arg96, which has a crucial role in the processive behavior [7], is flexible and able to bind (negatively charged) oxygen-containing molecules in several well-defined conformations. This may reflect the role of Arg96 in binding the polygalacturonic acid substrate, preventing its release, but at the same time being flexible enough to guide the substrate towards the active site.

Other factors besides Arg96 must contribute to the processivity of polygalacturonase I, since the Arg96Ser mutant of endopolygalacturonase I does not show the same product progression profile as wild-type endopolygalacturonase II and likewise, the Ser91Arg mutant of polygalacturonase II does not show the same product progression profile as wild-type endopolygalacturonase I [7]. Our structure suggests that the asparagine 126 insertion may be a good candidate as an additional contributor to the enzyme's processivity.

Acknowledgements: We are indebted to The European Commission, Europectin Project QLK3-1999-00089 for support. We thank the staff of the protein crystallography beam lines at the ESRF, Grenoble for the data collection facilities and assistance.

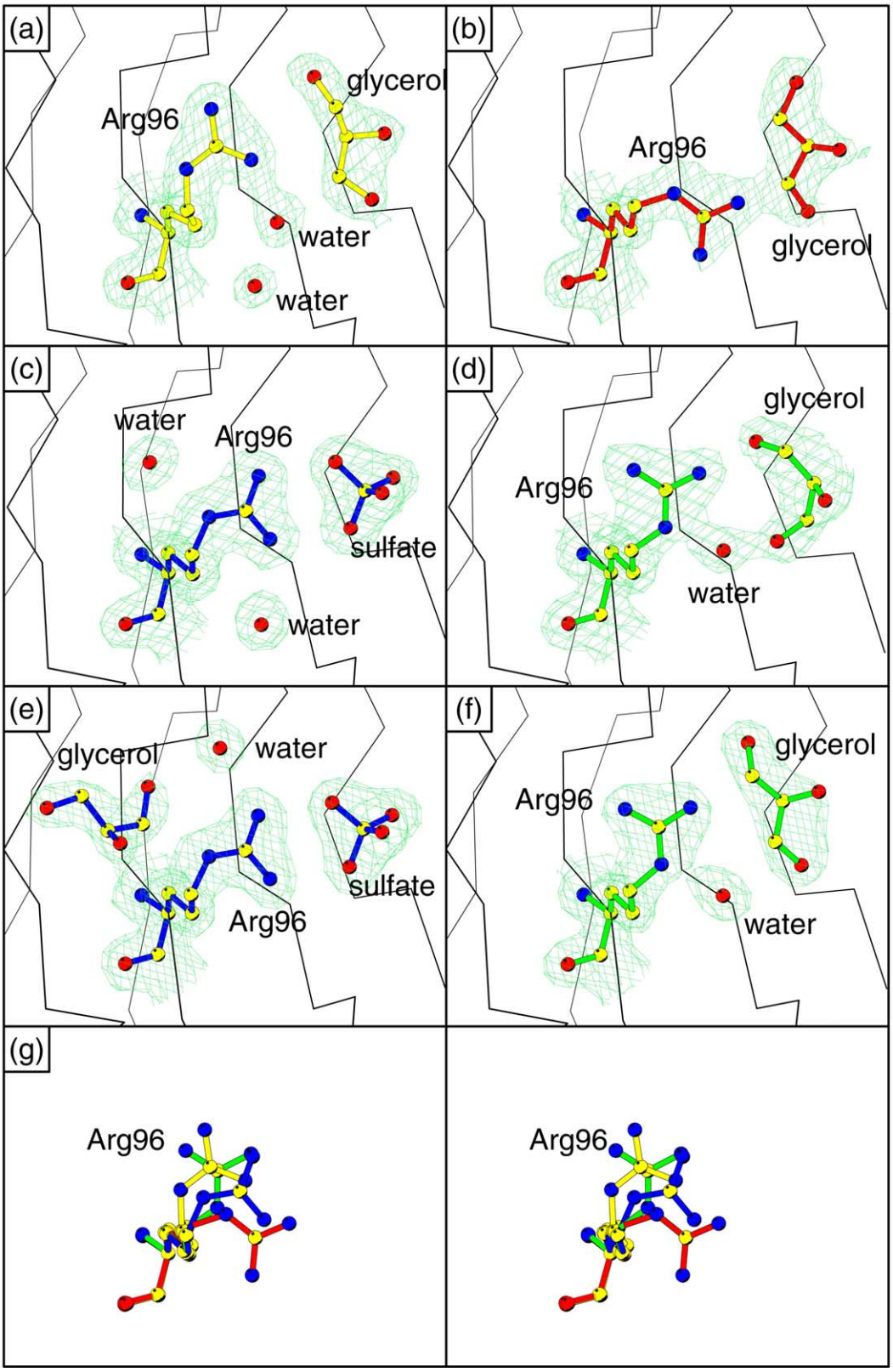

Fig. 3. A–F: The $2F_o-F_c$ electron density maps (contoured at 1σ) around Arg96 in the different molecules in the asymmetric unit showing the different well-defined conformations of this side chain. Arg96 and the interacting molecules are shown in ball-and-stick representation. The Cα trace is shown in black lines. G: A stereo view of the four most different conformations of Arg96 (A–D) superimposed [35].

References

[1] Parenicová, L., Benen, J.A.E., Kester, H.C.M. and Visser, J. (1998) Eur. J. Biochem. 251, 72–80.
[2] Benen, J.A.E., Kester, H.C.M. and Visser, J. (1999) Eur. J. Biochem. 259, 577–585.
[3] Pagès, S., Heijne, W.H.M., Kester, H.C.M., Visser, J. and Benen, J.A.E. (2000) J. Biol. Chem. 275, 29348–29353.
[4] Parenicová, L., Benen, J.A.E., Kester, H.C.M. and Visser, J. (2000) Biochem. J. 345, 637–644.
[5] Henrissat, B. (1991) Biochem. J. 280, 309–316.
[6] Henrissat, B. and Bairoch, A. (1993) Biochem. J. 293, 781–788.
[7] Pagès, S., Kester, H.C.M., Visser, J. and Benen, J.A.E. (2001) J. Biol. Chem. 276, 33652–33656.
[8] Armand, S., Wagemaker, M.J.M., Sánches-Torres, P., Kester, H.C.M., Van Santen, Y., Dijkstra, B.W., Visser, J. and Benen, J.A.E. (2000) J. Biol. Chem. 275, 691–696.
[9] Otwinowski, Z. and Minor, W. (1997) Methods Enzymol. 276, 307–326.
[10] French, S. and Wilson, K. (1978) Acta Crystallogr. A34, 517–524.
[11] Kelley, L.A., MacCallum, R.M. and Sternberg, M.J.E. (2000) J. Mol. Biol. 299, 499–520.
[12] Van Santen, Y., Benen, J.A.E., Schröter, K.-H., Kalk, K.H., Armand, S., Visser, J. and Dijkstra, B.W. (1999) J. Biol. Chem. 274, 30474–30480.
[13] Kissinger, C.R., Gehlhaar, D.K. and Fogel, D.B. (1999) Acta Crystallogr. D55, 484–491.
[14] Brünger, A. et al. (1998) Acta Crystallogr. D54, 905–921.
[15] Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) Acta Crystallogr. A47, 110–119.
[16] Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Acta Crystallogr. D53, 240–255.
[17] Holm, L. and Sander, C. (1996) Science 273, 595–602.
[18] Cho, S.W., Lee, S. and Shin, W. (2001) J. Mol. Biol. 314, 863–878.
[19] Shimizu, T., Nakatsu, T., Miyairi, K., Okuno, T. and Kato, H. (2002) Biochemistry 41, 9951–9959.
[20] Petersen, T.N., Kauppinen, S. and Larsen, S. (1997) Structure 5, 533–544.
[21] Pickersgill, R., Smith, D., Worboys, K. and Jenkins, J. (1998) J. Biol. Chem. 273, 24660–24664.
[22] Federici, L., Caprari, C., Mattei, B., Savino, C., Di Matteo, A., De Lorenzo, G., Cervone, F. and Tsernoglou, D. (2001) Proc. Natl. Acad. Sci. USA 98, 13425–13430.
[23] Van Asselt, E.J., Thunnissen, A.-M.W.H. and Dijkstra, B.W. (1999) J. Mol. Biol. 291, 877–898.
[24] Hol, W.G., van Duijnen, P.T. and Berendsen, H.J. (1978) Nature 273, 443–446.
[25] Schmidt, A., Schlacher, A., Steiner, W., Schwab, H. and Kratky, C. (1998) Protein Sci. 7, 2081–2088.
[26] Van Asselt, E.J., Kalk, K.H. and Dijkstra, B.W. (2000) Biochemistry 39, 1924–1934.
[27] Burmeister, W.P., Cottaz, S., Rollin, P., Vasella, A. and Henrissat, B. (2000) J. Biol. Chem. 275, 39385–39393.
[28] Mark, B.L., Vocadlo, D.J., Knapp, P., Triggs-Raine, B.L., Withers, S.G. and James, M.N.G. (2001) J. Biol. Chem. 276, 10330–10337.
[29] Parsiegla, G., Juy, M., Reverbel-Leroy, C., Tardif, C., Belaïch, J.-P., Driguez, H. and Haser, R. (1998) EMBO J. 17, 5551–5562.
[30] Parsiegla, G., Reverbel-Leroy, C., Tardif, C., Belaich, J.P., Driguez, H. and Haser, R. (2000) Biochemistry 39, 11238–11246.
[31] Breyer, W.A. and Matthews, B.W. (2001) Protein Sci. 10, 1699–1711.
[32] Zou, J. et al. (1999) Struct. Fold. Des. 7, 1035–1045.
[33] MacKenzie, L.F., Sulzenbacher, G., Divne, C., Jones, T.A., Wöldlike, H.F., Schülein, M., Withers, S.G. and Davies, G.J. (1998) Biochem. J. 335, 409–416.
[34] Kraulis, P.J. (1991) J. Appl. Crystallogr. 24, 946–950.
[35] Esnouf, R.M. (1997) J. Mol. Graph. Model. 15, 132–134.