The Three-dimensional Structures of Two β-Agarases*

Received for publication, July 30, 2003, and in revised form, September 8, 2003
Published, JBC Papers in Press, September 11, 2003 DOI 10.1074/jbc.M308313200

Julie Allouch‡, Murielle Jam§, William Helbert§, Tristan Barbeyron§, Bernard Kloareg§, Bernard Henriisset, and Mirjam Czjzek‡

From the ¤Laboratoire de Biochimie de Lignocellulose, UMR 6980, Université de Rouen and Laboratoire de Biochimie de Lignocellulose, Université de Caen, France
§Oligosaccharides, Défense et Développement Chez les Végétaux, UMR 1931, Centre de la Recherche Scientifique et Laboratoires Goémar, Station Biologique, BP 74, F-29682 Roscoff Cedex, France

Agars are important gelifying agents for biochemical use and the food industry. To cleave the β-1,4-linkages between β-D-galactose and α-L-3,6-anhydro-galactose residues in the red algal galactans known as agars, marine bacteria produce polysaccharide hydrolases called β-agarases. β-Agarases A and B from Zobellia galactanivorans Dsij have recently been biochemically characterized. Here we report the first crystal structure of these two β-agarases. The two proteins were overproduced in Escherichia coli and crystallized, and the crystal structures were determined at 1.48 and 2.3 Å for β-agarases A and B, respectively. The structure of β-agarase A was solved by the multiple anomalous diffraction method, whereas β-agarase B was solved with molecular replacement using β-agarase A as model. Their structures adopt a jelly roll fold with a deep active site channel harboring the catalytic machinery, namely the nucleophilic residues Glu-147 and Glu-184 and the acid/base residues Glu-152 and Glu-189 for β-agarases A and B, respectively. The structures of the agarases were compared with those of two lichenases and of a κ-carrageenase, which all belong to family 16 of the glycoside hydrolases in order to pinpoint the residues responsible for their widely differing substrate specificity. The relationship between structure and enzymatic activity of the two β-agarases from Z. galactanivorans Dsij was studied by analysis of the degradation products starting with different oligosaccharides. The combination of the structural and biochemical results allowed the determination of the number of subsites present in the catalytic cleft of the β-agarases.

Agarose is a hydrophilic polysaccharide found in the cell wall of marine red algae (Rhodophyceae) (1), where it naturally occurs in the form of a pseudocrystalline matrix associated with cellulose (2). It consists of a linear backbone of galactopyranose residues linked by alternating α-1,3- and β-1,4-linkages. Whereas the β-linked residues are in the β configuration, the α-1,3-linked galactose units are in the rare β configuration and are further modified by a 3,6-anhydro bridge (2, 3) (Fig. 1a). On the basis of x-ray fiber diffraction, optical rotation calculations, and solution gel transition, both a parallel double helix as well as a single helix structure for agarose have been proposed, where the individual polysaccharide chains have a left-handed 3-fold helix symmetry and a pitch of 1.90 nm (3, 4). Agarose forms thermo-reversible gels structured by aggregates of agarose chains. These gels exhibit unique rheological properties and are widely used as texturing agents for various applications in the food industry or as a common laboratory medium for chromatographic separation or bacterial colony growth (5).

Agarases are the glycoside hydrolases (GH) that hydrolyze agarose. The β-agarases and the α-agarases cleave the internal β-1,4- and α-1,3-linkage of agarose, respectively. β-Agarases produce agar-oligosaccharides in the series homologous to neoagarobiose (O-3,6-anhydro-α-1,galactopyranosyl-1,3-D-galactose), i.e. with 3,6-anhydro-D-galactose residues at the non-reducing ends and D-galactose residues at the reducing ends (6). β-Agarases are found in three families of the sequence-based classification of glycoside hydrolases (afmb.cnrs-mrs.fr/CAZY) (7–9), namely families GH-16, GH-50, and GH-86. Family GH-16 is the most abundant with β-agarases from Streptomyces coelicolor A3 (2, 10) (GenBank™ accession number X05811), from the pathogenic bacterium Aeromonas sp. (GenBank™ accession number U61972), from the marine bacteria Pseudomonas sp. ND137 (GenBank™ accession number AB063259), from Pseudoalteromonas sp. CY24 (GenBank™ accession number AY150179), from Pseudalteromonas atlantica ATCC 19262 (11, 12) (GenBank™ accession number MT3783), and from Zobellia galactanivorans Dsij (GenBank™ accession number AX008608 and AX008610 for β-agarase A and β-agarase B, respectively). The β-agarases I and II of P. atlantica are endo-enzymes acting randomly along the agarose chains (11, 12). Family GH-50 contains two β-agarases from Vibrio sp. JT0107 (13, 14) (GenBank™ accession number D14721 and D21202), whereas family GH-86 contains a β-agarase from P. atlantica (15) (GenBank™ accession number M22725). In contrast to those several known and somewhat characterized β-agarases, there is only one α-agarase reported at the sequence level (GenBank™ accession number AY164641), and it does not show any similarity to the β-agarases nor to any known GH family.

Although the sequence-based GH families contain enzymes of similar fold and identical molecular mechanism (16–18), they are known to often group together enzymes of varying substrate specificity (7–9). Interestingly, family GH-16 is such

* This work was supported by the region “Provence-Alpes-Côte d’Azur” (France), Goémar, St. Malo (France), and the CNRS (France).
† To whom correspondence should be addressed. Tel.: 33-491164513; Fax: 33-491164536; E-mail: czjzek@afmb.cnrs-mrs.fr.

1 The abbreviations used are: GH, glycoside hydrolases; MAD, multiple anomalous diffraction; CM, catalytic modules; HPAEC, by high performance anion exchange chromatography; DP, degree of polymerization.
a “polyspecific” family and contains enzymes hydrolyzing widely different substrates such as keratan-sulfate, β-1,3-glucan, mixed linkage β-1,3-(4)-glucans, xylanoglucom, κ-carrageenan, and agarose. These enzymes share a common ancestor and have diverged significantly in their primary sequence (17, 19). Pairwise comparisons of the sequences of family GH-16 reveals the existence of several subfamilies that coincide with the experimentally determined substrate specificity of the enzymes. Indeed five such subfamilies can be identified: (i) β-agarases, (ii) endo-1,3-β-glucanases (laminarinases) (iii), endo-β-1,3-1,4-glucanases (lichenases), (iv) κ-carrageenases, and (v) xylanoglucom endotransferases. Several additional subfamilies can be detected, but the substrate specificity of their members is not known at present. The few endo-β-galactosidases that have been characterized have too few homologues to allow the building of any subfamily. Although the overall sequence identity between members of a subfamily is typically better than 30–35%, inter-subfamily sequence identity is low (10–25%), and the similarity is apparently restricted to a few invariant residues that include the catalytic machinery. As a consequence of this conservation, all the glycoside hydrolases of family GH-16, including the two Z. galactanivorans Dsij β-agarases (β-agarases A and B), proceed with a molecular mechanism leading to overall retention of the anomeric configuration (20) and are hence capable of transglycosylating activity (18). This activity is actually the dominant feature of the family GH-16 enzymes that act on xylanoglucom (xylanoglucom endotransferases). The low degree of sequence similarity observed between the subfamilies within family GH-16 prevents successful homology modeling of an enzyme based on the structure of an enzyme with a different substrate specificity. Instead, and despite a predictable overall fold similarity, structural determinations are required to understand the details of substrate recognition and specificity in each subfamily. Three-dimensional structures have already been solved for the β-1,3-1,4-glucanase (20–23) and for the κ-carrageenase (24) subfamilies. Structural data representative of xylanoglucom endotransferases will soon be available, because the crystallization of this type of enzyme from Populus tremula × tremuloides has been reported recently (25). Because there is currently no three-dimensional structure available for any agarase, we have started a program aiming at the structural resolution of the two Z. galactanivorans Dsij β-agarases. The structural comparison between both enzymes as well as with the other GH-16 enzymes with known structures is expected to help define the factors governing substrate specificity within this class of enzymes.

Sequence analysis shows that the Z. galactanivorans Dsij β-agarase A gene (agaA; GenBankTM accession number AX008608) encodes for a modular protein comprising a signal peptide of 19 amino acids, the GH-16 catalytic module (residues 20–295), and a C-terminal domain (residues 296–539) of unknown function. In contrast to β-agarase A, the β-agarase B gene (agaB; GenBankTM accession number AX008610) encodes a single module protein consisting of 334 amino acids (residues 19–353). The first 18 amino acids are supposed to be a signal peptide of type II where residue Cys-18 might be linked to a lipid moiety (Fig. 1b). To avoid the known problems associated with the crystallization of multimodular proteins, we expressed both by the MAD method at 1.48 Å resolution and by molecular replacement at 2.3 Å resolution, respectively. These three-dimensional structures help to define the amino acid
residues involved in the recognition and cleavage of agarose and to understand the determinants of substrate specificity by comparison with the known structures of *Pseudoalteromonas carrageenovora* κ-carrageenase and the *Bacillus licheniformis* and *Paenibacillus macerans* β-1,3-1,4-glucanases. In addition, the number of sugar-binding subsites in both β-agarases was determined biochemically by high performance anion exchange chromatography analysis of the reaction products using different starting oligosaccharides.

**EXPERIMENTAL PROCEDURES**

**Protein Expression and Purification—**Escherichia coli ORIGAMI (DE3) pLysS strain cells harboring *Z. galactanivorans* Daji agaA and agaB genes in an expression plasmid pET20b (Novagen) have been described by Flament et al. Briefly, β-AgaA_CM and β-AgaB were expressed using the pET20b vector (Novagen) as a His-tagged fusion protein in the periplasm of *E. coli* ORIGAMI (DE3) pLysS strain and purified first by metal affinity chromatography on a column of Chelating Fast Flow-Sepharose loaded with NiSO₄ by using a concentration gradient from 30 to 400 mM imidazole. A second purification step by size exclusion chromatography on S100-Ni-Sephacryl (Amersham Biosciences) by using 20 mM Tris, pH 7.5, 200 mM NaCl yielded an electrophoretically pure protein (data not shown). The yield was 2 mg/liter culture medium, and the purified protein was concentrated using a dialyzing concentrator equipped with a polyether-sulfone membrane (10 kDa cut-off) (Amicon).

**Crystal Structure of β-Agarases**

2 Cells were grown in M9 medium supplemented with 2% casamino acids, 100 μg/ml ampicillin, 15 μg/ml tetracycline, 34 μg/ml chloramphenicol, and 15 μg/ml kanamycin.

Full details of the expression and purification of native β-AgaA_CM and β-AgaB have also been reported by Flament et al. Briefly, β-AgaA_CM and β-AgaB were expressed using the pET20b vector (Novagen) as a His-tagged fusion protein in the periplasm of *E. coli* ORIGAMI (DE3) pLysS strain and purified first by metal affinity chromatography on a column of Chelating Fast Flow-Sepharose loaded with NiSO₄ by using a concentration gradient from 30 to 400 mM imidazole. A second purification step by size exclusion chromatography on S100-Ni-Sephacryl (Amersham Biosciences) by using 20 mM Tris, pH 7.5, 200 mM NaCl yielded an electrophoretically pure protein (data not shown). The yield was 2 mg/liter culture medium, and the purified protein was concentrated using a dialyzing concentrator equipped with a polyether-sulfone membrane (10 kDa cut-off) (Amicon).

**Crystallization—**β-AgaA_CM and β-AgaB were concentrated to 2.5 mg/ml and stored in 50 mM Tris buffer, pH 7.5, with 25 mM NaCl. Crystallization conditions were first investigated using two sparse-matrix sampling kits (Molecular Dimensions and Stura Footprint).

The crystallization solutions contained 28% PEG 8000, 200 mM ammonium sulfate, and 100 mM sodium acetate, pH 4.6, for β-AgaA_CM and 58% methyl pentanediol, 20 mM calcium chloride, and 100 mM Hepes, pH 7.5, for β-AgaB. For both enzymes, crystals were grown by mixing 2 volumes of protein with 1 volume of precipitant solution in a hanging-drop vapor diffusion set up at 20.5 °C. Crystals grew within 1 week. Thereafter, a single crystal was soaked for 30 s in successive

### Table I

| Data statistics | β-agaA_CM (ID14-eh1) | β-agaB (ID14-eh2) |
|-----------------|----------------------|-------------------|
| **Wavelength (Å)** | 1.38530 | 1.38561 |
| | 1.38561 | 1.3482 |
| **f₀, fᵢ** | -17.6 | -18.4 |
| | -18.4 | -19.3 |
| **Redundancy** | 4.0 (4.1) | 4.0 (4.1) |
| | 4.0 (4.1) | 4.0 (4.1) |
| **Completeness (%)** | 99.5 (99.5) | 99.9 (99.9) |
| | 99.9 (99.9) | 99.7 (99.7) |
| **I/σI** | 162.1 (10.3) | 156.6 (9.2) |
| | 156.6 (9.2) | 156.6 (9.2) |
| **Rsym (%)** | 3.9 (5.8) | 3.7 (6.6) |
| | 3.7 (6.6) | 3.9 (7.6) |
| **Phasing (SOLVE)** | 15.1 (21.4) | 11.9 (19.8) |
| | 11.9 (19.8) | 12.3 (19.5) |
| **FOM before/after DM (to 2.5 Å)** | 0.39 (0.53) | |
| | |
| **Native data statistics** | β-agaA_CM | β-agaB |
| **Resolution (Å)** | 14.8-1.48 | 14.8-2.3 |
| | 14.8-2.3 | 14.8-2.3 |
| **Total data** | 454,198 | 428,617 |
| | 428,617 | 428,617 |
| **Unique data** | 42,255 | 64,783 |
| | 64,783 | 64,783 |
| **Completeness (%)** | 9.4 (13.9) | 9.4 (13.7) |
| | 9.4 (13.7) | 9.4 (13.7) |
| **I/σI** | 8.3 (37.4) | 10.6 (33.9) |
| | 10.6 (33.9) | 10.6 (33.9) |
| **Rsym (%)** | 3.9 (3.7) | 3.5 (3.4) |
| | 3.5 (3.4) | 3.5 (3.4) |
| **Completeness (%)** | 99.7 (99.7) | 99.9 (99.9) |
| | 99.9 (99.9) | 99.9 (99.9) |
| **I/σI** | 5.7 (1.9) | 5.4 (1.6) |
| | 5.4 (1.6) | 5.4 (1.6) |
| **Rsym (%)** | 8.3 (37.4) | 10.6 (33.9) |
| | 10.6 (33.9) | 10.6 (33.9) |
| **Quality of Ramachandran plot** | β-agaA_CM | β-agaB |
| **Percentage of residues in most allowed regions** | 87.4 | 87.1 |
| | 87.1 | 87.1 |
| **Percentage of residues in additional allowed regions** | 12.6 | 12.9 |
| | 12.9 | 12.9 |
| **Percentage of residues in generously allowed regions** | 0.0 | 0.0 |
| | 0.0 | 0.0 |
| **No. atoms** | β-agaA_CM | β-agaB |
| | 2182 | 18,742 |
| | 18,742 | 18,742 |
| **Water** | 366 | 410 |
| | 410 | 410 |
| **Ions (Na)** | 1 | 4 |
| | 4 | 4 |
| **Ions (Ca)** | 1 | 0 |
| | 0 | 0 |
| **Ions (SO₄)** | 3 | 0 |
| | 0 | 0 |
| **Ions (Hepe)** | 0 | 0 |
| | 0 | 0 |

*Values in parentheses correspond to the highest resolution shell.

a *R* = \[ \sqrt{\text{sum of squares of observed intensities} - \text{sum of squares of calculated intensity}} / \text{sum of squares of observed intensities} \] where the summation is over all symmetry equivalent reflections.

b \[ R_{\text{sym}} = \left[ \sum | I_\text{obs} - I_\text{calc} | / \sum I_\text{obs} \right] \] for anomalous differences, where *I* is the average of Friedel amplitudes at a single wavelength.

c \[ R_{\text{sym}} = \left[ \sum | I_\text{obs} - I_\text{calc} | / \sum | I_\text{obs} | \right] \] for dispersive differences, where *I* is the average amplitude at two wavelengths.

d FOM, figure of merit.

e \[ R \] calculated on 5% of data excluded from refinement.

f r.m.s.d., root mean square deviation.
solutions in which the glycerol concentration was increased by steps of 5% to a final concentration of 10% for β-AgaA_CM. For β-AgaB, the crystal was soaked 30 s in a 5% glycerol solution. All crystals were mounted on a loop, transferred to the goniometer head, and kept at 100 K in a nitrogen stream.

Data Collection and Phasing—Diffraction data at resolutions of 1.48 and 2.3 Å were collected on beamlines ID14-EH1 and EH2 (ESRF, Grenoble, France) for β-AgaA_CM and β-AgaB, respectively. For β-AgaA_CM, the space group was identified by autoindexing as P2₁ 2₁ 2₁, with cell constants a = 50.4 Å, b = 56.6, and c = 88.5 Å, suggesting a Vₘ₉ value of 2.0 Å³/Da with one molecule in the asymmetric unit (26). The data were processed with DENZO (27).

The structure was solved by the multiple anomalous diffraction (MAD) method (28) using ytterbium as anomalous diffuser. Ytterbium ions were incorporated by soaking with a mother solution containing 0.33 mM ytterbium chloride (exposure time 1 h and 15 min). The crystal was then back-soaked in the above-mentioned cryo-solution. The data were collected at three wavelengths around the LIII absorption edge of ytterbium (beamline ID29, ESRF, Grenoble, France). An ADSC image plate detector was used. All intensity data were integrated and reduced.

**Fig. 3.** a, stereo view of β-AgaA_CM (top) and β-AgaB (bottom) ribbon models. The calcium ion is displayed in purple, located on the convex side of the protein. The figure was produced with Molscript (44). b, electrostatic surface potential of β-AgaA_CM (left) and β-AgaB (right). Blue patches represent positive potential, red represent negative potential, and white surface is neutral. The number of subsites and their location within the molecule are represented by numbers (−4 to +4). The aromatic residues supposed to interact with the substrate are labeled at their location on the protein. The figure was produced with GRASP (45).
using MOSFLM/SCALa (29). The position of one ytterbium atom and the phase problem were solved using SOLVE with an overall Z score of 4.2 (30). The experimental phases from SOLVE resulted in a figure of merit of 0.67 at 2.5 Å resolution.

The electron density was improved by solvent flattening with the electron density map was of high quality (Fig. 2). The model was then refined at 1.48 Å resolution to a crystallographic R factor of 14.9% and R\text{free} factor of 17.6%. The three-dimensional crystal structure of \( \beta \)-AgaB was determined at 2.3 Å by molecular replacement with AMoRe using \( \beta \)-AgaA\_CM as the search model and was then refined to a crystallographic R factor of 19.9% and R\text{free} factor of 22.4%. Crystallographic statistics for both \( \beta \)-agarases are reported in Table I. In \( \beta \)-AgaA\_CM, the asymmetric unit encompasses amino acids 20–289. Residue Ser-290 is not visible in the (2\( F^o \) – \( F^c \)) electron density map and is thus presumed to be in a disordered conformation. The high resolution electron density allowed the detection of a sequence discrepancy (Alu-S4 is in fact a proline). The triplet CCG coding for a proline differs by a single nucleotide from the triplet GCC leading to an alanine, which easily accounts for an error that might have occurred during the sequence determination. In \( \beta \)-AgaB, the asymmetric unit encompasses amino acids 59–353. Residues 19–58 are not visible in the (2\( F^o \) – \( F^c \)) electron density map and are probably either truncated or in a disordered conformation.

As expected, the overall fold of both agarases is similar to the other members of family GH-16 with known structures (20–22, 24). Both \( \beta \)-agarases fold into globular ellipsoids, 58 × 47 × 42 Å and 56 × 47 × 40 Å in dimensions for \( \beta \)-AgaA\_CM and \( \beta \)-AgaB, respectively, and consist almost exclusively of \( \beta \)-sheets and surface loops (Fig. 3a). Each sheet is formed by seven \( \beta \)-strands. \( \beta \)-AgaB has three supplementary loops compared with \( \beta \)-AgaA\_CM; two of these loops are located on the same side, connecting strands at the C-terminal extremity of the internal \( \beta \)-sheet, and are close in space; the third is on the opposite side connecting strands at the N-terminal end of the external \( \beta \)-sheet. The molecules have a slightly elongated shape with a prominent cleft crossing one side. This cleft originates from the twist and bending of the two \( \beta \)-sheets, arranged on top of each other in a sandwich-like manner, and from the loop regions as already described for lichenase (37). The cleft is at least 35 Å long for both \( \beta \)-AgaA\_CM and \( \beta \)-AgaB and is equally lined with aromatic residues, listed in Table II.

**Structural Comparison with Family GH-16 Enzymes**—Fig. 4 shows the amino acid sequence alignment of \( \beta \)-AgaA\_CM and \( \beta \)-AgaB with the \( \kappa \)-carrageenase of \( P. \) carrageenovora (PDB code 1DYP) and the \( \beta \)-1,3–1,4-glucanases of \( B. \) licheniformis (PDB code 1GBG) and \( P. \) macrocerus (PDB code 1MAC). Despite the low sequence similarity of agarases with other members of family GH-16, most of the amino acids comprising the basic secondary structures of the \( \beta \)-jelly roll motif are well conserved. Consequently, the core region of the \( \beta \)-sandwich is perfectly superimposable, reflected by the lower root mean square deviations for the C\text{a} atoms of the \( \beta \)-sandwich. The overall root mean square deviations of all structurally known members of family GH-16 enzymes are summarized in Table III.

The sandwich fold of \( \beta \)-AgaA\_CM and \( \beta \)-AgaB has the same number of \( \beta \)-strands as in \( \kappa \)-carrageenase (24) and in \( \beta \)-1,3–
1,4-glucanases (37). However, some extra structural elements are found in the loops connecting the internal and external strands of the two β-sheets. For instance, helix H9252 and strand H9252 are located in loops bordering the substrate binding cleft on the side where the reducing end of the polysaccharide would be.

In the two β-agarases and in the κ-carrageenase, strands H9252 and H9252 are separated by a short loop insertion. In the β-1,3-1,4-glucanases, this insertion is missing, yielding a single long β-strand. On the side of the cleft, where the non-reducing end of the substrate would bind, a short helix (α1) stabilizes a loop insertion in β-AgaA_CM and β-AgaB, closing the end of the cleft. These structural differences at the border of the substrate-binding site are most probably responsible for the differences in substrate specificity of the enzymes. At this point, it is interesting to note that β-agarases A and B are the only members of family GH-16 that recognize and cleave 1,4-glycosidic units linked to L-3,6-anhydro-galactose moieties, whereas the κ-carrageenase cleaves 1,4-glycosidic units linked to D-3,6-anhydro-galactose moieties.
Table III
Root mean square deviations in Angstrom (boldface values) and number of Cα of the superpositioning of the Ca traces

|              | β-AgaA_CM (269 Cα) | β-AgaB (296 Cα) | Core sandwich of β-agarase_CM (110 Cα) |
|--------------|--------------------|-----------------|-------------------------------------|
| β-AgaB       | 269 Cα             | 296 Cα          | 109 Cα                              |
| k-Carrageenase | 223 Cα             | 228 Cα          | 104 Cα                              |
| Lichenase BM | 206 Cα             | 203 Cα          | 94 Cα                               |
| Lichenase BL | 204 Cα             | 203 Cα          | 93 Cα                               |

Anhydrosialylase. This should have a global consequence on the environment of the binding cleft, because the 3,6-anhydro-galactose units whether in the l- or d-galacto configuration are much more bulky than the analogous units, and consequently staking interactions are only possible on one face of the sugar unit. The largest differences with β-1,3-1,4-2 glucanases are therefore expected in the -2 and -4 subsites on the glycine-binding side and at the +1 and +3 aglycone-binding subsites, following the subsite naming convention of Davies et al. (38). Indeed, the inserted loops mentioned above in both agarases and the k-carrageenase give different forms to the cleft at the level of the extremities of the binding subsites, which appear to be less linear in the cases of the β-agarases (Fig. 3b). Differences between agarases and the other members of family GH-16 at the −1 subsite should only concern the O-4 position and are discussed below under “Substrate-binding Region.” Two short helices (a3 and a4), only present in the agarases, are located in loops on the side of the external β-sheet and are therefore not responsible for differences in substrate specificity.

In β-AgaA_CM, a calcium ion is bound with nearly perfect octahedral geometry on the convex face of the protein, coordinating to the backbone carboxyl oxygen atoms of Ser-47, Ser-91, and Asp-279, a carboxylate oxygen of Asp-279, one of Asn-49, and one of Asp-22, and one water molecule. In β-AgaB, a calcium ion is located at the same position but coordinated in a pentahedral manner to the backbone carboxyl oxygen atom of Asn-83, Gly-127, and Asp-343, a carboxylate oxygen of Asp-343, and one water molecule. It is interesting to note that this structural calcium ion is also found in the other GH-16 members with known three-dimensional structures (lichenases and k-carrageenase) (24, 37) and that the aspartic acid responsible for the only side-chain interaction with calcium is conserved (Fig. 4).

Catalytic Machinery—Glycosidic bond cleavage by retaining glycosidases such as those in family GH-16 takes place via a two-step mechanism involving two enzymatic carboxyl groups (39–41). In the first step (glycosylation), one of the carboxyl group acts first as a general acid, which protonates the glycosidic oxygen and thereby facilitates aglycone departure. Simultaneously, the other catalytic residue acts as a nucleophile, attacking the anomeric carbon and forming a covalent glycosyl-enzyme intermediate. In the second step (deglycosylation), the general acid now acts as a general base and deprotonates a water molecule, in order to hydrolyze the glycosyl-enzyme intermediate and to release the product. Both steps occur via transition states with substantial oxocarbenium ion character (40, 41). In family GH-16 glycoside hydrolases, the catalytic nucleophile and acid/base residues have been unambiguously identified in the P. macerans and B. licheniformis 1,3–1,4-β-glucanases (21, 22) and have been confirmed in the P. carrageenovora k-carrageenase (24). Sequence and structure similarities between members of family GH-16 allowed the rapid identification of the catalytic machinery of the two β-agarases on strand β9 with Glu-147 and Glu-184 as the nucleophile and Glu-152 and Glu-189 as the acid/base in β-AgaA_CM and β-AgaB, respectively. These two residues are found in the deepest region of the cleft of both enzymes. A third conserved acidic residue (Asp-149 and Asp-186, respectively, in β-AgaA_CM and β-AgaB) points toward the cleft and is probably important to maintain the charges in the environment of the catalytic amino-acids.

Like the k-carrageenase, the two β-agarases have a hydrophobic residue inserted between the two catalytic residues on strand β9 (Val-166, Ile-150, and Ile-188 in k-carrageenase, β-AgaA_CM, and β-AgaB, respectively). This is in marked contrast to the β-1,3–1,4-glucanases. Interestingly, a β-bulge in the agarases and the k-carrageenase (24) “compensates” for this insertion, and in consequence all the enzymes have a strictly superimposable catalytic machinery (Fig. 5). As common for retaining glycoside hydrolases, the carboxyl groups of the nucleophile Glu-147 and Glu-184 and the acid/base catalyst Glu-152 and Glu-189 amino acids are separated by 6.2 and 5.8 Å for β-AgaA_CM and β-AgaB, respectively.

In k-carrageenase, oxygen OD2 of Asp-165 makes a 2.74-Å long hydrogen bond to His-183. Michel et al. (24), have suggested that Asp-165 and His-183 might be implicated in proton traffic, governing the regeneration of the protonated states of the catalytic residues during catalysis and cooperating in proton trafficking during the deglycosylation step. This histidine is conserved in both agarases; however, the positioning is not the same as in the k-carrageenase. In β-AgaB, ND1 of His-215 is at 3.5 Å from Asp-186, whereas the high resolution (1.48 Å) of β-AgaA_CM shows clearly that the side-chain of His-172 has an opposite orientation, and its ND1 atom is hydrogen-bonded to OD1 of Asp-181 (2.68 Å) and not to Asp-149. In the lichenases this histidine residue is replaced by a tyrosine, which is not involved in hydrogen-bonding with the catalytic residues.

Substrate-binding Region—In the absence of a complex with a substrate molecule, it is difficult to analyze the elements governing substrate binding. No complex is available either with the other GH-16 enzymes with known three-dimensional structures. However, it has been shown that family GH-16 and family GH-7 are structurally and mechanistically related despite the absence of detectable sequence similarity outside of a common sequence motif, E(ILV)D(IVAF)(VILMF)(0,1)E. The
structural similarity between the two families, accompanied by a perfect correspondence of the catalytic machinery, has allowed their grouping in a superfamily called “clan GH-B” (22, 24). We have taken advantage of the availability of the three-dimensional structure of *Fusarium oxysporum* endoglucanase I of family GH-7 complexed with a nonhydrolysable substrate analogue (PDB code 1OVW) (42) in order to infer by similarity the approximate location and direction of the agarose substrate in the active site of *β*-agarases. After superimposition of the catalytic residues and of the secondary structure elements, the cleft on the concave side of both *β*-agarases A and B has a position similar to that of *F. oxysporum* endoglucanase (42). From this superimposition, some differences between the family GH-16 members may be highlighted at the level of subsites -1 and +1. As mentioned above, the only difference in the *β*-linked pyranose unit that binds the -1 subsite of the licheneses, *κ*-carrageenase, and *β*-agarases is the O-4 position. In the substrate of licheneses, this unit is a glucose residue where the OH-4 group is equatorial. In contrast this group is axial in the galactose unit of agarose and axial and sulfated in *κ*-carrageenan. Indeed, corresponding differences are seen with respect to the residues of the different enzymes that are located in the area involved in potential hydrogen bonds to O-4. In both *β*-agarases a glutamic acid (308 in *β*-AgaB and 254 in *β*-AgaA_CM) and a glutamine (310 in *β*-AgaB and 256 in *β*-AgaA_CM) are candidates for hydrogen bonding with the axial O-4 group of the sugar unit in subsite -1. In the *κ*-carrageenase these residues are replaced by Gly-258 and Arg-260, the latter clearly being in favor of binding to a sulfate group, whereas the former is probably necessary to create room to accommodate the sulfate group. In the licheneses, the glutamic acid (Glu-254 in *β*-AgaA_CM) is replaced by a tryptophan residue, which could form a hydrogen bond to the equatorial O-4 of the glucose unit that comes to lie in this subsite. In the place of Gln-256 of *β*-AgaA_CM (Gln-310 in *β*-AgaB) the licheneses possess a phenylalanine (Phe-30 in *B. licheniformis* lichenase (1GBG)), which comes from a completely different structural element than in both the agarases or the *κ*-carrageenase. Interestingly, the alignment of licheneses and xyloglucan endotransfereases, which are all members of family GH-16 that degrade polysaccharides having the OH-4 group in equatorial position, reveal that this tryptophan is strictly conserved within these members. The position of Phe-30 in *B. licheniformis* lichenase is less conserved but is most often a tyrosine or a phenylalanine and in all cases a hydrophobic residue in most licheneses and xyloglucan endotransfereases. As expected, the

![Detailed view of selected residues in the active site cleft of family GH-16 enzymes. a, β-AgaA_CM; b, β-AgaB; c, κ-carrageenase; d, lichenase from B. licheniformis. The catalytic glutamates are colored in red; important residues for substrate binding in subsite -1 are colored in indigo; putative residues involved in substrate specificity are shown in yellow; a single amino acid difference between AgaA_CM and AgaB, close to the catalytic site, is colored in pink. The figure was prepared with Molscript (44).](http://www.jbc.org/)

**Fig. 6.** Detailed view of selected residues in the active site cleft of family GH-16 enzymes. a, β-AgaA_CM; b, β-AgaB; c, κ-carrageenase; d, lichenase from *B. licheniformis*. The catalytic glutamates are colored in red; important residues for substrate binding in subsite -1 are colored in indigo; putative residues involved in substrate specificity are shown in yellow; a single amino acid difference between AgaA_CM and AgaB, close to the catalytic site, is colored in pink. The figure was prepared with Molscript (44).
Crystal Structure of β-Agarases

that it may contain eight subsites. In order to determine the number of subsites in the two agarases, we have carried out an analysis of the reaction products of β-agarases A and B by using oligosaccharides of defined degree of polymerization as substrates and high performance anion exchange chromatography detection. The degradation products observed after incubation of calibrated oligo-agaroses either with β-agarase A or B were similar. The apparent rate of hydrolysis was found to be similar when the concentration of β-AgaA_CM was 50 times higher than that of β-AgaB (Fig. 7a). This lower specific activity of β-AgaA_CM toward oligosaccharides was more pronounced in the case of the hexa-acetate neoagarohexose. When this oligosaccharide was subjected to enzymatic assays, the amount of neoagarotetraose detected was about three times lower for β-AgaA_CM than for β-AgaB (Fig. 7b). Neoagaroctaose was fragmented only in tetrasaccharides, whereas neoagarodecaose was disrupted in tetra- and hexasaccharides (Fig. 7c). This degradation pattern indicates that the active site can accommodate a maximum of eight sugar units, in agreement with the structural analysis above. Both enzymes degrade agarose oligosaccharides comprising at least six sugars to the same final length. However, β-AgaA_CM preferably degrades longer oligosaccharides and is overall less efficient than β-AgaB. The difference between both enzymes is kinetic and not in the release of different reaction products. We have found only a single amino acid difference between the two enzymes in the close vicinity of the catalytic residues. At subsite −1, the aspartic acid 173 in β-AgaA is replaced by Ala-136 in β-AgaA_CM. It is conceivable that this change could alter slightly the pH optimum of the general acid/base, thereby explaining the subtle difference in catalytic activity between the two enzymes.

Our structural and biochemical data clearly show that β-agarases A and B from Z. galactanivorans Dsi2 are endo-agarases, having a long and open active site cleft that accommodates at least eight sugar units. The comparison with the other members in family GH-16 with known structures allowed us to identify the residues that are most probably involved in substrate specificity. More precise details of agarose recognition obviously require the production of a catalytically inactive mutant and the co-crystallization with a substrate molecule. This work is currently underway in our laboratories.

Acknowledgments—We thank Gurvan Michel for many helpful discussions and Audrey Lartigue for assistance during protein expression and purification. We are also indebted to the ESRF staff on beamlines ID14-EH1, EH2, and ID29 for help during data collections.

REFERENCES

1. Knoll, A. H. (1992) Science 256, 622–627
2. Klaaregg, B., and Quatrano, R. S. (1988) Oceanog. Mar. Biol. Annu. Rev. 26, 259–315
3. Arnot, S., Fulmer, A., Scott, W. E., Dea, I. C., Moorhouse, R., and Rees, D. A. (1974) J. Mol. Biol. 90, 269–284
4. Rochas, C., Lahaye, M., Yaphé, W., and Phan Viet, M. T. (1986) Carbohydr. Res. 145, 199–207
5. De Ruyter, G. A., and Rudolph, B. (1979) Trends Food Sci. Technol. 8, 389–395
6. Malot, C., Jimenez-Barbero, J., Bernabe, M., Brosa, C., and Planas, A. (1991) Biochem. J. 286, 753–758
7. Henriissat, B. (1991) Biochem. J. 280, 309–318
8. Henriissat, B., and Bairoch, A. (1993) Biochem. J. 283, 781–788
9. Henriissat, B., and Bairoch, A. (1993) Biochem. J. 231, 605–606
10. Buttner, M. J., Fearnley, I. M., and Bish, M. J. (1987) Mol. Gen. Genet. 101, 101–109
11. Morriss, M. L., McLean, M. W., Long, W. F., and Williamson, F. B. (1983) Eur. J. Biochem. 137, 149–154
12. Morriss, M. L., McLean, M. W., Williamson, F. B., and Long, W. F. (1983) Eur. J. Biochem. 135, 553–558
13. Sugano, Y., Matsumoto, T., Kodama, H., and Noma, M. (1993) Appl. Environ. Microbiol. 59, 3750–3756
14. Sugano, Y., Matsumoto, T., and Noma, M. (1994) Biochim. Biophys. Acta 1218, 105–108
15. Belas, R. (1989) J. Bacteriol. 171, 602–605
16. Gehler, J., Gilkes, N. R., Claysens, M., Wilson, D. B., Béguin, P., Wackarchuk, W. W., Kilburn, D. G., Miller, R. C., Jr., Warren, R. A. J., and Withers, S. G. (1992) J. Biol. Chem. 267, 12529–12561
17. Davies, G., and Henrissat, B. (1995) Structure 3, 853–859
Crystal Structure of β-Agarases

47180

18. Henrissat, B., and Davies, G. (1997) Curr. Opin. Struct. Biol. 7, 637–644
19. Barbeyron, T., Gerard, A., Petin, P., Henrissat, B., and Kloareg, B. (1998) Mol. Biol. Evol. 15, 528–537
20. Keitel, T., Simon, O., Borriss, R., and Heinemann, U. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5287–5291
21. Hahn, M., Olsen, O., Politz, O., Borriss, R., and Heinemann, U. (1995) J. Biol. Chem. 270, 3081–3088
22. Juncoea, M., Pons, J., Dot, T., Querol, E., and Planas, A. (1994) J. Biol. Chem. 269, 14530–14535
23. Li-Chu, T., Lie-Fen, S., Shu-Hua, L., Su-Shiang, L., and Hann, S. Y. (2003) J. Mol. Biol. 330, 607–620
24. Michel, G., Chantalat, L., Duee, E., Barbeyron, T., Henrissat, B., Kloareg, B., and Dideberg, O. (2001) Acta Crystallogr. Sect. D Biol. Crystallogr. 57, 513–525
25. Johansson, P., Denman, S., Brumer, H., Kallas, A. M., Henriksson, H., Bergfors, T., Teeri, T. T., and Jones, T. A. (2003) Acta Crystallogr. Sect. D Biol. Crystallogr. 59, 535–537
26. Matthews, B. W. (1968) J. Mol. Biol. 33, 491–497
27. Otwinowski, Z., and Minor, W. (1997) Methods Enzymol. 276, 307–326
28. Terwilliger, T. C., and Berendzen, J. (1997) Acta Crystallogr. Sect. D Biol. Crystallogr. 53, 571–579
29. Powell, H. B. (1999) Acta Crystallogr. Sect. D Biol. Crystallogr. 55, 1690–1695
30. de La Fortelle, E., and Bricegne, G. (1997) Methods Enzymol. 276, 472–494
31. CCP4 (1994) Acta Crystallogr. Sect. D Biol. Crystallogr. 50, 760–763
32. Navaza, J. (1994) Acta Crystallogr. Sect. A. 50, 157–161
33. Roussel, A., and Cambillau, C. (1991) Silicon Graphics Geometry Partners Directory 88, Mountain View, CA
34. Perrakis, A., Sixma, T. K., Wilson, K. S., and Lamzin, V. S. (1997) Acta Crystallogr. Sect. D Biol. Crystallogr. 53, 448–455
35. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) J. Appl. Crystallogr. 26, 283–291
36. Weinberger, F., Richard, C., Kloareg, B., Kashman, Y., Hoppe, H.-G., and Friedlander, M. (2001) J. Physiol. 37, 418–426
37. Hahn, M., Pons, J., Planas, A., Querol, E., and Heinemann, U. (1995) FEBS Lett. 374, 221–224
38. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) Biochem. J. 321, 557–559
39. Koshland, D. E. (1953) Biol. Rev. Camb. Philos. Soc. 28, 416–436
40. Sinnott, M. L. (1990) Chem. Rev. 90, 1171–1202
41. McCarter, J. D., and Withers, S. G. (1994) Curr. Opin. Struct. Biol. 4, 885–892
42. Poulos, G. G., Driguez, H., Henrissat, B., Schulein, M., and Davies, G. J. (1996) J. Bacteriol. 178, 15280–15287
43. Czjzek, M., Ciesi, M., Zamboni, V., Bevan, D. R., Henrissat, B., and Essen, A. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 13555–13560
44. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–950
45. Nicholls, A., Sharp, K., and Hong, B. (1993) Proteins 11, 281–296
46. Gut, P., Courcelle, E., Stuart, D. I., and Metz, F. (1999) Bioinformatics 15, 305–308
The Three-dimensional Structures of Two β-Agarases
Julie Allouch, Murielle Jam, William Helbert, Tristan Barbeyron, Bernard Kloareg,
Bernard Henrissat and Mirjam Czjzek

J. Biol. Chem. 2003, 278:47171-47180.
doi: 10.1074/jbc.M308313200 originally published online September 11, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M308313200

Alerts:
  • When this article is cited
  • When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 38 references, 8 of which can be accessed free at
http://www.jbc.org/content/278/47/47171.full.html#ref-list-1