Family 6 Carbohydrate Binding Modules in β-Agarases Display Exquisite Selectivity for the Non-reducing Termini of Agarose Chains**

Joanna Henshaw†, Ami Horne-Bitschy‡, Alicia Lammerts van Bueren∥, Victoria A. Money‡, David N. Bolam‡, Mirjam Czjzek‡, Nathan A. Ekborg**, Ronald M. Weiner**, Steven W. Hutcheson**, Gideon J. Davies§, Alisdair B. Boraston§†, and Harry J. Gilbert**‡

The polysaccharides in the cell wall of both marine and terrestrial plants represent the most abundant reservoir of organic carbon in the biosphere. The microbial hydrolysis of these polymers is not only central to the carbon cycle but also of considerable industrial significance. Indeed, the enzymes that catalyze this process are already widely used in the biosphere. The microbial hydrolysis of these polymers is not only central to the carbon cycle but also of considerable industrial significance.

Carbohydrate recognition is central to the biological and industrial exploitation of plant structural polysaccharides. These insoluble polymers are recalcitrant to microbial degradation, and enzymes that catalyze this process generally contain non-catalytic carbohydrate binding modules (CBMs) that potentiate activity by increasing substrate binding. Agarose, a repeat of the disaccharide 3,6-anhydro-L-galactose-(1,3)-β-D-galactopyranose-(1,4), is the dominant matrix polysaccharide in marine algae, yet the role of CBMs in the hydrolysis of this important polymer has not previously been explored. Here we show that family 6 CBMs, present in two different β-agarases, bind specifically to the non-reducing end of agarose chains, recognizing only the first repeat of the disaccharide. The crystal structure of one of these modules Aga16B-CBM6-2, in complex with neoagarohexaose, reveals the mechanism by which the protein displays exquisite specificity, targeting the equatorial O4 and the axial O3 of the anhydro-L-galactose. Targeting of the CBM6 to the non-reducing end of agarose chains may direct the appended catalytic modules to areas of the plant cell wall attacked by β-agarases where the matrix polysaccharide is likely to be more amenable to further enzymic hydrolysis.

The polysaccharides in the cell wall of both marine and terrestrial plants represent the most abundant reservoir of organic carbon in the biosphere. The microbial hydrolysis of these polymers is not only central to the carbon cycle but also of considerable industrial significance. Indeed, the enzymes that catalyze this process are already widely used in the biosphere. The microbial hydrolysis of these polymers is not only central to the carbon cycle but also of considerable industrial significance.

Carbohydrate recognition is central to the biological and industrial exploitation of plant structural polysaccharides. These insoluble polymers are recalcitrant to microbial degradation, and enzymes that catalyze this process generally contain non-catalytic carbohydrate binding modules (CBMs) that potentiate activity by increasing substrate binding. Agarose, a repeat of the disaccharide 3,6-anhydro-L-galactose-(1,3)-β-D-galactopyranose-(1,4), is the dominant matrix polysaccharide in marine algae, yet the role of CBMs in the hydrolysis of this important polymer has not previously been explored. Here we show that family 6 CBMs, present in two different β-agarases, bind specifically to the non-reducing end of agarose chains, recognizing only the first repeat of the disaccharide. The crystal structure of one of these modules Aga16B-CBM6-2, in complex with neoagarohexaose, reveals the mechanism by which the protein displays exquisite specificity, targeting the equatorial O4 and the axial O3 of the anhydro-L-galactose. Targeting of the CBM6 to the non-reducing end of agarose chains may direct the appended catalytic modules to areas of the plant cell wall attacked by β-agarases where the matrix polysaccharide is likely to be more amenable to further enzymic hydrolysis.

The polysaccharides in the cell wall of both marine and terrestrial plants represent the most abundant reservoir of organic carbon in the biosphere. The microbial hydrolysis of these polymers is not only central to the carbon cycle but also of considerable industrial significance. Indeed, the enzymes that catalyze this process are already widely used in the biosphere. The microbial hydrolysis of these polymers is not only central to the carbon cycle but also of considerable industrial significance.

Carbohydrate recognition is central to the biological and industrial exploitation of plant structural polysaccharides. These insoluble polymers are recalcitrant to microbial degradation, and enzymes that catalyze this process generally contain non-catalytic carbohydrate binding modules (CBMs) that potentiate activity by increasing substrate binding. Agarose, a repeat of the disaccharide 3,6-anhydro-L-galactose-(1,3)-β-D-galactopyranose-(1,4), is the dominant matrix polysaccharide in marine algae, yet the role of CBMs in the hydrolysis of this important polymer has not previously been explored. Here we show that family 6 CBMs, present in two different β-agarases, bind specifically to the non-reducing end of agarose chains, recognizing only the first repeat of the disaccharide. The crystal structure of one of these modules Aga16B-CBM6-2, in complex with neoagarohexaose, reveals the mechanism by which the protein displays exquisite specificity, targeting the equatorial O4 and the axial O3 of the anhydro-L-galactose. Targeting of the CBM6 to the non-reducing end of agarose chains may direct the appended catalytic modules to areas of the plant cell wall attacked by β-agarases where the matrix polysaccharide is likely to be more amenable to further enzymic hydrolysis.

The polysaccharides in the cell wall of both marine and terrestrial plants represent the most abundant reservoir of organic carbon in the biosphere. The microbial hydrolysis of these polymers is not only central to the carbon cycle but also of considerable industrial significance. Indeed, the enzymes that catalyze this process are already widely used in the biosphere. The microbial hydrolysis of these polymers is not only central to the carbon cycle but also of considerable industrial significance.

Carbohydrate recognition is central to the biological and industrial exploitation of plant structural polysaccharides. These insoluble polymers are recalcitrant to microbial degradation, and enzymes that catalyze this process generally contain non-catalytic carbohydrate binding modules (CBMs) that potentiate activity by increasing substrate binding. Agarose, a repeat of the disaccharide 3,6-anhydro-L-galactose-(1,3)-β-D-galactopyranose-(1,4), is the dominant matrix polysaccharide in marine algae, yet the role of CBMs in the hydrolysis of this important polymer has not previously been explored. Here we show that family 6 CBMs, present in two different β-agarases, bind specifically to the non-reducing end of agarose chains, recognizing only the first repeat of the disaccharide. The crystal structure of one of these modules Aga16B-CBM6-2, in complex with neoagarohexaose, reveals the mechanism by which the protein displays exquisite specificity, targeting the equatorial O4 and the axial O3 of the anhydro-L-galactose. Targeting of the CBM6 to the non-reducing end of agarose chains may direct the appended catalytic modules to areas of the plant cell wall attacked by β-agarases where the matrix polysaccharide is likely to be more amenable to further enzymic hydrolysis.

The polysaccharides in the cell wall of both marine and terrestrial plants represent the most abundant reservoir of organic carbon in the biosphere. The microbial hydrolysis of these polymers is not only central to the carbon cycle but also of considerable industrial significance. Indeed, the enzymes that catalyze this process are already widely used in the biosphere. The microbial hydrolysis of these polymers is not only central to the carbon cycle but also of considerable industrial significance.

Carbohydrate Recognition is central to the biological and industrial exploitation of plant structural polysaccharides. These insoluble polymers are recalcitrant to microbial degradation, and enzymes that catalyze this process generally contain non-catalytic carbohydrate binding modules (CBMs) that potentiate activity by increasing substrate binding. Agarose, a repeat of the disaccharide 3,6-anhydro-L-galactose-(1,3)-β-D-galactopyranose-(1,4), is the dominant matrix polysaccharide in marine algae, yet the role of CBMs in the hydrolysis of this important polymer has not previously been explored. Here we show that family 6 CBMs, present in two different β-agarases, bind specifically to the non-reducing end of agarose chains, recognizing only the first repeat of the disaccharide. The crystal structure of one of these modules Aga16B-CBM6-2, in complex with neoagarohexaose, reveals the mechanism by which the protein displays exquisite specificity, targeting the equatorial O4 and the axial O3 of the anhydro-L-galactose. Targeting of the CBM6 to the non-reducing end of agarose chains may direct the appended catalytic modules to areas of the plant cell wall attacked by β-agarases where the matrix polysaccharide is likely to be more amenable to further enzymic hydrolysis.

The polysaccharides in the cell wall of both marine and terrestrial plants represent the most abundant reservoir of organic carbon in the biosphere. The microbial hydrolysis of these polymers is not only central to the carbon cycle but also of considerable industrial significance. Indeed, the enzymes that catalyze this process are already widely used in the biosphere. The microbial hydrolysis of these polymers is not only central to the carbon cycle but also of considerable industrial significance.

Carbohydrate Recognition is central to the biological and industrial exploitation of plant structural polysaccharides. These insoluble polymers are recalcitrant to microbial degradation, and enzymes that catalyze this process generally contain non-catalytic carbohydrate binding modules (CBMs) that potentiate activity by increasing substrate binding. Agarose, a repeat of the disaccharide 3,6-anhydro-L-galactose-(1,3)-β-D-galactopyranose-(1,4), is the dominant matrix polysaccharide in marine algae, yet the role of CBMs in the hydrolysis of this important polymer has not previously been explored. Here we show that family 6 CBMs, present in two different β-agarases, bind specifically to the non-reducing end of agarose chains, recognizing only the first repeat of the disaccharide. The crystal structure of one of these modules Aga16B-CBM6-2, in complex with neoagarohexaose, reveals the mechanism by which the protein displays exquisite specificity, targeting the equatorial O4 and the axial O3 of the anhydro-L-galactose. Targeting of the CBM6 to the non-reducing end of agarose chains may direct the appended catalytic modules to areas of the plant cell wall attacked by β-agarases where the matrix polysaccharide is likely to be more amenable to further enzymic hydrolysis.
CBM6 in β-Agarases Selects for Non-reducing Agarose

Although the ligand binds to the concave β-sheet in the majority of CBM families with this fold, the site of saccharide recognition varies in different members of CBM6. In the laminarase- and xylanase-derived CBM6s the ligand binding site is located within the loops connecting the two β-sheets (20, 21), termed “cleft A.” By contrast, the concave surface “cleft B” of CmlLic-CBM6 binds to internal regions of cellulose chains and β-1,4-β-1,3 linked glucans, whereas the non-reducing termini of xylan and cellulose are accommodated within the loops that form cleft A (19, 22). Thus, not only does the ligand binding site vary in CBM6, but the family contains “type B” (for example, Cxyl-CBM6 and Cxyl-CBM6) binding sites, while CmlCBM6–2 displays both type B and type C ligand binding modes.

Agarose is an important marine polysaccharide of molecular weight ~120,000 (24), based on the repeating unit 3,6-anhydro-α-L-galactose-(1,3)-β-D-galactopyranose-(1,4) (Fig. 1) (25). The polysaccharide forms a double helical paracrystalline structure, which is highly recalcitrant to enzyme attack (26). The agarose backbone is hydrolyzed by α- and β-agarases, which cleave the α-1,3- and β-1,4-glycosidic bonds, respectively (25). The vast majority of agarases described in the literature act on β-glycosidic bonds (β-agarases), whereas there are only three enzymes (all in glycoside hydrolase family 96) in the CAZY data base (afmb.cnrs-mrs.fr/CAZY/) that are classified as β-agarases. The mechanism by which these enzymes are able to access single agarose chains is currently unclear, although the recent description of the crystal structure of a β-agarase reveals an auxiliary agarose binding site in the catalytic module that may contribute to local unwinding of the double helical structure (25). Although several CBM6s are located in β-agarases, their functional significance is unclear. Here we show that three family 6 CBMs located in two Saccharophagus degradans β-agarases, Aga16B and Aga86E (so called because they contain catalytic modules belonging to glycoside hydrolase families 16 and 86, respectively), bind to the non-reducing end of single agarose chains. The crystal structure of one of these CBMs in complex with neogalactosaccharides reveals that, despite a paucity of hydrogen bonds between ligand and protein, the exploitation of the hydroxyls that are unique to the 3,6-anhydro-α-L-galactose-(1,3)-β-D-galactopyranose-(1,4) repeating unit of agarose as recognition determinants contributes to the exquisite specificity displayed by these protein modules for the marine polysaccharide. The functional significance of the binding mode of these CBM6s is discussed.

MATERIALS AND METHODS

Protein Expression and Purification—The DNA sequences encoding the Aga16B and Aga86E family 6 CBMs were amplified by PCR from plasmids containing aga16B and aga86E, respectively, using the thermostable DNA polymerases Pfu (Invitrogen, amplified aga16B) or Pfu Turbo (Stratagene, amplified aga86E) and the primers listed in Table S1 (supplemental material). The amplified DNA from aga86E was cloned into NdeI- and XhoI-restricted pET22b to generate pH1 (Aga86E-CBM6-1), pH2 (Aga86E-CBM6-2), and pH3 (Aga86E-CBM6-3). Amplified DNA encoding the CBM6 modules of Aga16B was ligated into NheI- and HindIII-digested pET28a to generate pH4 (Aga16B-CBM6-1) and pH5 (Aga16B-CBM6-2), respectively (Fig. 2). The encoded recombinant proteins, which contain a C-terminal His6 tag, were expressed in Escherichia coli strains TUNER (Novagen, Aga86E-derived CBMs) or BL21 STAR DE3 (Aga16B-derived CBMs), harboring the appropriate recombinant plasmid. Cultures were incubated at 37 °C to mid exponential phase (A600 ~0.7), isopropyl-β-D-thiogalactopyranoside was added to a final concentration of 200 μM, and the cells were incubated for a further 16 h at 16 °C. The bacteria were harvested and fractionated to produce cell-free extract and insoluble fractions. The His6-tagged recombinant proteins in the insoluble fraction were redissolved in 20 mM Tris/HCl buffer containing 500 mM NaCl (Buffer A) and 6 M urea and purified by immobilized metal ion affinity chromatography. The bound protein was refolded on the immobilized nickel metal ion column by washing the matrix with a step gradient of 6 to 0 M urea in Buffer A. The refolded protein was eluted using a 0 to 300 mM imidazole gradient in Buffer A. To generate selenomethionine (SeMet) Aga16B-CBM6-2 the E. coli methionine auxotroph B834(DE3) was cultured as described in Carvalho et al. (27), and the protein was purified using the same procedures employed for the native CBMs, except 2 mM β-mercaptoethanol was included in all buffers. For crystalization experiments Aga16B-CBM6-2 was concentrated using an Amicon 10-kDa molecular mass centrifugal concentrator and washed three times in 5 mM dithiothreitol (for the SeMet proteins) or water (for the native proteins).

Generation of CBM6 Mutants—Mutants of Aga16B-CBM6-2 were generated using the PCR-based QuikChange kit (Stratagene) according to the manufacturer’s instructions, employing pH5 as the template DNA and the primers listed in Table 1S. To investigate the ligand-binding site of the protein, the mutant W97M was generated. To assist in solving the crystal structure of the protein, five methionine mutants (I5M, A65M, V69M, L77M, and L131M) were produced, because the native CBM6 does not contain any methionine residues. These mutants were combined in various permutations, and we used the ISM/L77M double mutant to solve the crystal structure of the protein.

Source of Sugars Used—Agarose, oat spelt xylan, laminarin, hydroxyethylcellulose, and neogalactosaccharides were obtained from Sigma, putulan was from Calbiochem, and cellooligosaccharides were from Seikagaku Corp. (Japan). All other sugars were purchased from Megazyme International (Bray, County Wicklow, Ireland). Reduced neogalactoheaxose was prepared by incubating the oligosaccharide with sodium borohydride as described previously (28).

Ligand Binding Studies—The qualitative binding of the CBM6 proteins to soluble polysaccharides was assessed by affinity gel electrophoresis, as described previously (29). To incorporate agarose, the polysaccharide (10% w/v) was solubilized by boiling and then added to the polycrylamide gels prior to polymerization. Ligand binding was also assessed by isothermal titration calorimetry. Isothermal titration calorimetry measurements were made at 25 °C following standard procedures (19) using a MicroCal Omega titration calorimeter. Proteins were dialyzed, extensively, against 50 mM sodium Hepes buffer, pH 7.0, containing 5 mM CaCl₂, and the ligand was dissolved in the same buffer to minimize heats of dilution. During a titration experiment the protein sample (40–250 μM), stirred at 300 rpm in a 1.4331-ml reaction cell maintained at 25 °C, was injected with a single 1-μl aliquot, followed by 29 successive 10-μl injections of oligosaccharide ligand (0.5–5 mM) at 200-s intervals. Integrated heat effects, after correction for heats of dilution, were analyzed by non-linear regression using a single-site binding model (MicroCal Origin, version 7.0). The fitted data yielded the association constant (Kₐ) and the enthalpy of binding (ΔH). Other thermo-
dynamic parameters were calculated using the standard thermodynamic equation: $-RT \ln K_s = \Delta G = \Delta H - T \Delta S$. To investigate the role of calcium in ligand binding the metal was removed with 10 mM EDTA.

Structure Solution and Crystallization of Aga16B-CBM6-2—Large single diffraction quality crystals of both the native protein and the I5M/L77M mutant were obtained by the hanging drop, vapor diffusion method with a mother liquor of 2M NaCl and 16–20% polyethylene glycol 4000 buffered to pH 7.5 with 100 mM Tris/HCl. The protein was used at concentration of 30 mg/ml and co-crystallized with an excess of 5m M) of neoagarohexaose, and drops were formed using 1 l. At 18 °C, were harvested in rayon fiber loops with cryoprotection provided by short soaks in mother liquor supplemented with 10–15% ethylene glycol before being flash frozen in liquid nitrogen.

Crystals of native Aga16BCBM6-2 co-crystallized with neoagarohexaose, and drops were formed using 1 micro liter of protein and 1 micro liter of well solution. The crystals, which formed overnight at 18 °C, were harvested in rayon fiber loops with cryoprotection provided by short soaks in mother liquor supplemented with 10–15% ethylene glycol before being flash frozen in liquid nitrogen.

Crystals of native Aga16BCBM6-2 co-crystallized with neoagarohexaose and soaked in cryoprotectant were mounted on the detector and frozen directly in the nitrogen stream at 113 K. The 1.6-Å resolution data were collected with a Rigaku R-AXIS 4++ area detector coupled to an MM-002 x-ray generator with Osmic “blue” optics. Data were processed using the Crystal Clear/d*trek software provided with the instrument (30). A subset of the observations (5%) was flagged as “free” and used to monitor refinement procedures (31). Data statistics are given in Table 1.

Table 1.

| Dataset | Aga16B-CBM6-2 native | Aga16B-CBM6-2 I5M/L77M mutant SeMet peak | Aga16B-CBM6-2 I5M/L77M mutant SeMet inflection | Aga16B-CBM6-2 I5M/L77M mutant SeMet remote |
|---------|----------------------|----------------------------------------|---------------------------------|----------------------------------|
| Beam line | N/A | ID-23 | ID-23 | ID-23 |
| Wavelength (Å) | 1.5417 | 0.97910 | 0.97565 | 0.97950 |
| Resolution of data (Å) | 19.62-1.61 | 96.12-1.6 (1.7-1.6)* | 65.64-2.1 (2.2-2.1) | 98.64-2.1 (2.2-2.1) |
| $R_{merge}$ | 0.047 (0.343) | 0.109 (0.140)* | 0.092 (0.108) | 0.108 (0.116) |
| Mean I/σI | 13.1 (2.9) | 16.9 (3.7)* | 20.6 (14.2) | 18.7 (13.9) |
| Completeness (%) | 98.5 (96.1) | 69.5 (11.3)* | 99.6 (97.5) | 99.5 (97.5) |
| Multiplicity | 3.4 (3.0) | 6.3 (3.1)* | 7.1 (6.6) | 7.1 (6.4) |

* The data were integrated into the corners of the detector to provide better resolution for data solution and automated structure building.

RESULTS

Production of Agarase-derived CBMs—The β-agarases Aga86E and Aga16B from S. degradans are modular enzymes containing a GH86 and GH16 catalytic module appended to three and two CBMs, respectively (Fig. 2). To explore the function of these CBMs, attempts were made to express these modules in E. coli. All of these CBMs could be produced as insoluble inclusion bodies. Of these all except Aga16B-CBM6-1 could be refolded by removal of urea (~70% solubilization) and purified to electrophoretic homogeneity by immobilized metal ion affinity chromatography.
CBM6 in β-Agarases Selects for Non-reducing Agarose

FIGURE 3. Isothermal titration calorimetry of Aga86E-CBM6-1 with neoagarobiose, neoagarotetraose, and neoagarohexaose. Aga86E-CBM6-1 was titrated with neoagarooligosaccharides in 50 mM sodium HEPES buffer, pH 7.0, with 5 mM CaCl$_2$ at 25 °C. The protein concentration of Aga86E-CBM6-1 for each titration was 100 μM.

TABLE 2

Binding of neoagarooligosaccharides to agarase-derived CBM6s as determined by isothermal titration calorimetry

| Protein          | Ligand               | $K_a$ ($10^{-5}$ M$^{-1}$) | $\Delta G$ (kcal mol$^{-1}$) | $\Delta H$ (kcal mol$^{-1}$) | $\Delta S$ (cal mol$^{-1}$ K$^{-1}$) | $n$  |
|------------------|----------------------|-----------------------------|------------------------------|-------------------------------|-----------------------------------|------|
| Aga86E-CBM6-1    | Neoagarobiose        | 10.1 ± 0.5                  | -6.8 ± 0.0                   | -14.6 ± 0.2                   | -7.8 ± 0.2                        | 0.9 ± 0.1 |
| Aga86E-CBM6-1    | Neoagarotetraose     | 15.0 ± 0.3                  | -7.1 ± 0.0                   | -11.9 ± 0.1                   | -4.8 ± 0.3                        | 1.1 ± 0.1 |
| Aga86E-CBM6-1    | Neoagarohexaose      | 15.2 ± 0.2                  | -7.1 ± 0.1                   | -12.4 ± 0.1                   | -5.3 ± 0.3                        | 0.9 ± 0.0 |
| Aga86E-CBM6-1    | Neoagarohexaose + 5 mM EDTA | NB$^b$                     | -7.1                       | -11.8                        | -4.7                              | 1.0 ± 1 |
| Aga86E-CBM6-1    | Neoagarohexaose reduced | 15.7 ± 0.1                  | -7.0 ± 0.0                   | -16.9 ± 0.4                   | -9.9 ± 0.3                        | 0.8 ± 0.3 |
| Aga86E-CBM6-2    | Neoagarobiose        | 12.6 ± 0.3                  | -7.0 ± 0.4                   | -15.9 ± 0.5                   | -8.9 ± 0.5                        | 0.8 ± 0.2 |
| Aga86E-CBM6-2    | Neoagarotetraose     | 13.0 ± 0.8                  | -7.0 ± 0.4                   | -12.8 ± 0.3                   | -5.8 ± 0.4                        | 0.9 ± 0.1 |
| Aga86E-CBM6-2    | Neoagarohexaose      | 13.5 ± 0.2                  | -7.0 ± 0.1                   | -10.3 ± 0.1                   | -3.7 ± 0.3                        | 1.1 ± 0.1 |
| Aga16B-CBM6-2    | Neoagarobiose        | 6.5 ± 0.5                   | -6.6 ± 0.1                   | -10.3 ± 0.1                   | -3.8 ± 0.3                        | 1.1 ± 0.1 |
| Aga16B-CBM6-2    | Neoagarotetraose     | 6.2 ± 0.5                   | -6.5 ± 0.1                   | -10.3 ± 0.1                   | -3.8 ± 0.3                        | 1.1 ± 0.1 |
| Aga16B-CBM6-2    | Neoagarohexaose + 5 mM EDTA | NB$^b$                     | -6.6 ± 0.4                  | -13.1 ± 0.5                   | -6.5 ± 0.5                        | 1.2 ± 0.0 |
| Aga16B-CBM6-2    | Neoagarohexaose + 5 mM EDTA | NB$^b$                     | -6.5 ± 0.0                  | -12.9 ± 0.0                   | -6.4 ± 0.1                        | 1.0 ± 0.1 |

$^a$ Binding was carried out in 50 mM sodium HEPES, pH 7.0, containing 5 mM CaCl$_2$ except when EDTA was included in the titration, in which case the divalent metal ion was omitted.

$^b$ NB, no binding detected.

Ligand Specificity of Agarase-derived CBM6s—The capacity of the CBM6s to bind to polysaccharides was explored by affinity gel electrophoresis. The data revealed no electrophoretic retardation of the three modules in the presence of soluble cellulose, xylans, mannanos, pectins, and agarose (data not shown). Despite its apparent inability to bind solubilized agarose, initial UV difference experiments (data not shown) suggested that Aga16B-CBM6-2 binds neoagarobiose in the presence of calcium, prompting further exploration of neoagarooligosaccharide binding by isothermal titration calorimetry. Example data are displayed in Fig. 3, and the full data set is presented in Table 2. Although Aga86E-CBM6-3 did not recognize these ligands, the other three modules bound to neoagarobiose, neoagarotetraose, and neoagarohexaose, which comprise one two and three repeats of the disaccharide repeating unit of agarose, 3,6-anhydro-α-L-galactose-(1,3)-β-D-galactopyranose-(1,4) (Fig. 1), respectively. D-Galactose and 3,6-anhydro-L-galactose did not interact with the CBM6s. The affinities of the three ligands for each of the CBM6s were similar, with a $K_a$ of $\sim$10$^5$. The experimental stoichiometries were $\sim$1 indicating that each of the refolded CBM6s contains a single ligand binding site. Oligosaccharide binding was enthalpically driven while the change in entropy had a detrimental influence on affinity. The thermodynamics of ligand binding displayed by the three modules is typical of the interaction of CBMs with soluble ligands, and the possible mechanisms for this loss in entropy and increase in enthalpy have been extensively discussed previously (38–41).

The observation that neoagarooligosaccharides with a degree of polymerization of 2, 4, or 6 displayed the same affinity for the CBM6s indicates that these protein modules contain two sugar-binding subsites. The apparent lack of binding to D-galactose and 3,6-anhydro-L-galactose suggests that tight binding only occurs when both subsites are occupied. The lack of binding to solubilized agarose as observed by affinity gel electrophoresis, which is an extremely sensitive method able to determine $K_a$ values as low as $10^{-5}$ M$^{-1}$ (29), is at odds with the tight binding to the repeating disaccharide unit. Binding to agarose should be observed by affinity gel electrophoresis if the large number of internal disaccharide units in this polysaccharide are indeed recognized by the CBM6s. That this was not the case suggests that the chain ends, which are relatively infrequent in a polymer with a degree of polymerization $\sim$1000, may be the recognition determinant for these CBMs. Aga86E-CBM6-1 bound equally well to untreated and sodium borohydride-reduced neoagarohexaose (sodium borohydride converts the reducing end sugar to the alditol causing ring opening (Table 2)) indicating that the reducing end is not a recognition determinant. The subsequent crystal structure of Aga16B-CBM6-2 in complex with neoagarohexaose confirms that at least this CBM is specific for 3,6-anhydro-α-L-galactose at the non-reducing end of agarose chains.

Crystal Structure of Aga16B-CBM6-2—Crystals of CBM6 were obtained in the presence of calcium and either neoagarobiose or neoagarohexaose but not in the absence of ligand. The crystals belong to the space group P2$_1$2$_1$2$_1$, with unit cell parameters of $a = 54.0$ Å, $b = 55.0$ Å, and $c = 196.9$ Å and four molecules in the asymmetric unit, each comprising residues 1–138, two ions, one neoagarohexaose molecule, and 140

17102  JOURNAL OF BIOLOGICAL CHEMISTRY  VOLUME 281 • NUMBER 25 • JUNE 23, 2006
water molecules. The coordinates have been deposited at the Protein Data Bank. Refinement statistics are given in Table 1.

Aga16B-CBM6-2 displays a typical “jelly roll” β-sandwich fold of two β-sheets (designated β-sheets 1 and 2). β-Sheet 1 contains β-strands 2 (Ala14–Gly16), 5 (Tyr46–Val53), 10 (Gly113–Gly121), 7 (Ser73–Asn79), and 8 (Trp84–Ala91) and is twisted displaying a concave surface, whereas β-sheet 2 consists of β-strands 1 (Ser4–Glu8), 11 (Asn130–Ser137), 6 (Gly57–Gly66), 9 (Gln101–Leu110), 4 (Thr35–Asn39), and 3 (Ser28–Thr32). The β-strands 4 and 3 are aligned at an angle of 70° with the other β-strands in β-sheet 2. The β-strands in both β-sheets are antiparallel (Fig. 4). The region extending from Gly17 to Pro26 is highly mobile and has only been modeled in one of the four molecules in the asymmetric unit. The overall fold of Aga16B-CBM6-2 is very similar to other CBM6 modules. This is confirmed by a DALI search (42), which shows that the protein displays closest structural similarity to two family 6 CBMs, followed by a CBM36 and several toxins. The root mean square deviations when superimposing Aga16B-CBM6-2 onto CmLic-CBM6 and CtxXyl-CBM6 are 1.7 Å (117 Ca) and 1.8 Å (107 Ca), respectively. Also common to this fold is a structural metal ion, which in CmCBM6–1 (23), CmCBM6 (22), and CtxCBM6 (20) is thought to be a calcium atom, whereas in BhLam-CBM6 it appears to be a sodium (21). This position is equivalent to the structural calcium ion found in several members of the jelly roll superfamily of CBMs, including CBM4 and CBM22 (17, 43, 44), where it contributes to stability (45). In Aga16B-CBM6-2 the electron density and presence of anomalous signal (not shown) make the identity of this metal ion consistent with a calcium atom (termed Ca1). This metal has a typical octahedral geometry making coordinate bonds with the O1 and backbone carbonyl of Asp154, a bidentate interaction with Glu32, O1 of Gly31, the carbonyl of Thr58, and a water molecule. This calcium ion can be considered as a structural ion bridging the N- and C-terminal region of Aga16B-CBM6-2.

Clear electron density was seen for a single neoagarohexaose moiety bound to each of the four Aga16B-CBM6-2 molecules in the asymmetric unit (Fig. 4). The CBM primarily interacts with the non-reducing disaccharide repeat of neoagarohexaose, consistent with the isothermal titration calorimetry data presented in Table 2. The bound sugar adopts an extended conformation with a bond angle around the glycosidic bond of 114°. This conformation is similar to the topology of neoagarooligosaccharides bound to β-agarase A from Zobellia galactanivorans (25) and the modeled structure of agarooligosaccharides in solution (46). The hexasaccharide interacts with the protein between the loops that connect the two β-sheets by linking β-strands 8 to 9, 4 to 5, and 10 to 11, respectively, which has been termed cleft A in other family 6 CBMs (20).
In the majority of CBM families that display a jelly roll fold the β-sheet that presents a concave surface (termed cleft B in Ref. 20 and equivalent to β-sheet 1 in Aga16B-CBM6-2) accommodates the target ligand (17, 27, 43, 44). By contrast the ligand binding site varies in different members of CBM6. Although family 6 CBMs derived from two xylanases and a laminarase bind to their respective ligands in cleft A, CmLig-CBM6 binds to internal regions of β-glucans in cleft B, while also interacting, albeit weakly, with terminal glucose and xylose residues in cleft A (19, 22). It is interesting to note, therefore, that while the first disaccharide repeat unit of the neoagaroehexose interacts with cleft A the ligand also occupies cleft B in adjacent molecules in the crystal lattice. The observations that the hexasaccharide displays the same affinity for Aga16B-CBM6 as the disaccharide, the stoichiometry of binding is one and the substitution of the cleft A residue Trp97 completely abrogates carbohydrate recognition indicate that the interaction with cleft B is a crystallographic “artifact.” Furthermore, Aga16B-CBM6-2 was also co-crystallized with excess neoagarobiose, and this ligand was only observed in cleft A (data not shown). Thus, we conclude that in Aga16B-CBM6-2 ligand binding in solution only occurs in cleft A.

Inspection of Aga16B-CBM6-2, in complex with neoagarohexose, reveals how the limited number of direct interactions between the protein and ligand leads to the tight specificity displayed by the agarase-derived module. Consistent with other family 6 CBMs two aromatic residues, Trp97 and Tyr40 make hydrophobic interactions with the non-reducing terminal sugar, 3,6-anhydro-α-D-galactose. Both aromatic side chains are orientated perpendicular to the sugar ring (Fig. 4), whereas the equivalent residues in other CBM6 modules are in a parallel orientation with the planar α and β face of the bound sugars. This is consistent with a greater distance between Trp97 and Tyr40 in Aga16B-CBM6-2 (9.5 Å) compared with other family 6 CBMs where the corresponding amino acids are separated by 8.5 Å (20, 22). Indeed, in all other CBMs characterized to date, aromatic residues, rather than aliphatic amino acids, are co-planar with the D-configured sugar rings with which they make extensive hydrophobic interactions (for review see Ref. 5). It is interesting to note, therefore, that in one of the few examples of a CBM binding to a L-sugar the aromatic residues are at 90° with the sugar rings. In this configuration, the equatorial hydrogens attached to C2, C3, and C5 of 3,6-anhydro-α-D-galactose will be pointing at the center of the aromatic rings of Trp97 and Tyr40 consistent with favorable “ring-current” interactions. This view is consistent with the observation that similar weak interactions are also likely to occur between the axial hydrogens at C2, C4, C4, and C5 of most D-sugars and the aromatic side chains of CBMs, which are aligned parallel to the pyranose rings of the cognate ligand. Similarly, in the complexes of the GH16 agarase 16AgaA from Z. galactanivorans Ds17 there are few aromatic interactions with the anhydro sugar, but on the single occasion they do occur they “stack” perpendicular to the sugar ring, as observed here (25).

“Classical” hydrogen bonds between Aga16B-CBM6-2 and its ligand are restricted to the 3,6-anhydro-α-L-galactose at the non-reducing terminus; the O4 of the sugar interacts with the N82 of both Asn130 and Asn39, respectively, and O3 makes a hydrogen bond with the backbone N of Tyr40 (Fig. 4). Although O3 and O4 of the anhydro sugar also make numerous indirect hydrogen bonds with the protein, the significance of these water-mediated interactions is unclear, because previous studies have shown that mutating amino acids in CBMs that make solvent-mediated hydrogen bonds with the target saccharides often has little influence on affinity (47, 48). Although O2 of 3,6-anhydro-α-L-galactose is within hydrogen bonding distance (3.5 Å) with the phenolic oxygen of Tyr40, the perpendicular orientation of the two atoms suggests that they do not make a significant interaction. Indeed the substitution of Tyr40 for phenylalanine in Aga86E-CBM6-2 (Fig. 5) does not influence the affinity of the protein for its ligand, confirming that hydrophobic interactions are the primary mechanism by which the phenolic side chain contributes to ligand binding. The sugar ring of the adjacent 6-D-galactose residue stacks against the side chain of Trp157 (Fig. 4), although the sugar makes no hydrogen bonds with the protein. The orientation of the hydrogen bonding network and the hydrophobic interactions directs the O4 of the terminal 3,6-anhydro-L-galactose at the surface of the protein explaining why Aga16B-CBM6-2 displays specificity for the non-reducing end of neoagarooligosaccharides. The interaction with the axial O3 and equatorial O4 of the terminal sugar confers specificity for anhydro-L-galactose (generated by the action of β-agarases) as opposed to other sugars, such as D-xylose, D-glucose, and D-mannose, which are ubiquitous in plant structural polysaccharides and have equatorial O3 atoms, or the D-galactose which would be at the non-reducing terminus of oligosaccharides produced by the action of α-agarases in which O3 is equatorial and O4 axial.

The residues in Aga16B-CBM6-2 that interact with neoagarohexose are invariant (Trp156, Asn130, Asn130, and Trp127) or highly conserved (the equivalent residue to Aga16B-CBM6-2 Tyr62 is Tyr40 in Aga86E-CBM6-1 and Phe27 in Aga86E-CBM6-2) in the other CBMs known to bind single agarose chains (Fig. 5). The loop extending from Thr18–Ser28 in Aga16B-CBM6-2 is deleted in Aga86E-CBM6-3, which does not bind to neoagarooligosaccharides. In the Aga16B CBM this loop contains Phe19, which makes a hydrophobic interaction with Tyr40, and plays a key role in positioning this residue within the ligand binding site (Fig. 4). The deleted loop also contains Asp21 that coordinates the calcium, which also interacts with Tyr40 (see below). The loss of these interactions in Aga86E-CBM6-3 will likely result in Tyr31 (equivalent to Aga16B-CBM6-2 Tyr40) adopting a different conformation, which may explain why the protein does not recognize neoagarohexose. Identification of the sequence motifs in the agarase CBMs that play a key role in the binding of these modules to the non-reducing ends of agarose chains informs the identification of other family 6 proteins that are likely to recognize the 3,6-anhydro-α-L-galactose-(1,3)-β-D-galactopyranose-(1,4) terminal disaccharide of the marine polysaccharide. There are 15 putative and confirmed β- and α-agarases in the Cazy database (GenBankTM accession codes are as follows: AF121273, BAE06228.1, BAC90022.1, BAD29947.1, BAD68321.1, AKK62837.1, AKK62838.1, BAD88713.1, AAT67062.1, AAP49346.1, AAP49316.1, AAP70390.1, AAP70365.1, AAQ53915.1, and AAQ53916.1); these contain a total of 21 CBM6s, which display sequence motifs from the corresponding Aga16B and Aga86E, the modules of which play a key role in binding the non-reducing end of agarose chains. By analogy to the Cellvibrio mixtus mixed-linked glucan binding CBM6, which binds via an endo mode to its ligand in cleft B, some agarase-derived CBM6s may recognize internal regions of agarose chains, and it is unlikely that the key saccharide binding amino acids in these modules will be conserved in Aga16B-CBM6-2, Aga86E-CBM6-1, and Aga86E-CBM6-2.

Role of Calcium in Ligand Binding—To investigate the role of calcium in ligand binding, the affinity of the three CBMs for neoagarohexose in the presence of 10 mM EDTA was assessed. The data (Table 2) revealed that the chelating agent caused the complete abrogation of ligand binding indicating that calcium (or some other divalent cation) plays a pivotal role in the recognition of neoagarooligosaccharides by the agarase-derived CBM6s.

Inspection of the location of the second metal in AgaB16-CBM6-2, interpreted as a solvent-exposed calcium, shows that it makes coordinate bonds with Asp21, Asn130, Asp120, Pro34, Tyr40, and a water molecule (Fig. 4). Although the metal does not make direct interactions with the
bound ligand, by contributing to the conformation adopted by Tyr40, which is integral to ligand recognition (discussed above), it plays an indirect role in the binding of neogarooligosaccharides to Aga16B-CBM6-2. Because the residues that coordinate the two calciums through their side chains are invariant in the three agarase CBMs that bind to neoagarooligosaccharides (Fig. 5), it is likely that the metal plays a similar role in ligand binding in all these proteins. The role of the metal ion in CBM6 is similar to that of calcium in legume lectins, which also contributes to the conformation adopted by the amino acids that interact with the carbohydrate ligand (for review see Ref. 49), although it does not interact directly with the bound sugar.

While metal ions do not play a key role in the binding of the majority of CBMs to their ligands, recent studies have shown that calcium ions are integral to carbohydrate recognition in two xylan-binding CBMs from families 35 and 36, respectively (50, 51). Although the metal plays a direct role in ligand recognition by the CBM36, it interacts with the protein and the O2 and O3 of a xylose in the binding site, the mechanism by which the calcium mediates xylan recognition by the family 35 CBM is currently unknown.

**DISCUSSION**

This report provides the first evidence that CBMs are able to bind to the non-reducing ends of agarose chains with specificity derived from interactions with hydroxyl groups that adopt conformations that are unique to 3,6-anhydro-α,1-galactose. Inspection of the cleft A ligand binding site of family 6 CBMs provides insight into the structural basis for the different specificities displayed by these protein modules. This site contains three highly conserved residues, Tyr40, Trp97, and Asn130 in Aga16B-CBM6-2, which contribute to ligand binding (Fig. 5), while additional binding sites are formed by the surrounding loops. These loops display significant differences between the different CBM6s and thus determine the variation in ligand specificity displayed by this family (Fig. 6). The laminarin binding CBM6 (BhLam-CBM6) has a five-residue insertion (loop 2 in Fig. 6, residues 124–129) in the loop closest to the C-terminal of the protein, around which the U-shaped saccharide is draped in the protein-ligand complex. Although the nature of the residues is not conserved, this loop insertion is also present in Aga16B-CBM6-2 and contains Trp127, which plays a pivotal role in the second
sugar binding site. This loop in both Aga16B–CBM6-2 and BhCBM6 adopts a similar conformation (Fig. 6), which occludes the cleft that binds to the internal regions of xylan chains in CtCBM6. Indeed the interaction between neogaro- hexose and Trp forces the ligand into a perpendicular orientation with respect to the surface of the protein, where extension of the non-reducing end of the oligosaccharide is prevented by steric clashes with β-strand 3, most notably with the side chain of Tyr (Fig. 6). Thus, it is primarily loop 2 in both BhCBM6 and Aga16B–CBM6-2 that confers the type C binding mode displayed by these protein modules, although the closure of the N-terminal side of the binding pocket by the phenolic ring of Tyr also contributes to the recognition of the non-reducing end of agarose chains by Aga16B–CBM6-2. A feature that is unique to the agarase CBM6 is an extensive loop insertion in the N-terminal region of the protein (loop 1 in Fig. 6, residues 20–28), which is involved in binding the second calcium. It is close to the ligand binding site and contributes indirectly to neogaro- lligosaccharide recognition (Fig. 4). This is in contrast to the CBM6s, which bind to glucose and xylose polymers, where residues from this smaller loop (Ile and Ser in CtCBM6; Glu in CmCBM6) contribute directly to ligand binding.

The presence of two CBMs in Aga86E (and probably also in Aga16B) is particularly intriguing. There are numerous examples of glycoside hydrolases that contain multiple CBMs, which display similar ligand specificities, and it is well established that these modules confer increased affinity through avidity affects (52, 53). It is likely, therefore, that the multiple agarase-binding CBM6s in a single protein, by interacting with agarose molecules that physically associate, mediate high affinity for these polymeric ligands via an avidity mechanism. It is also possible that the multiple CBM6s in Aga16B and Aga86E bind to both chains in a single double helical agarose molecule. This could maintain the ends of the two strands in an open conformation such that they are not able to physically associate, thus increasing their accessibility to the catalytic modules of these enzymes. Indeed, CBM20s that target amylose, which also displays a double helical structure, contain two similar but distinct starch binding sites, and there are some biochemical and microscopic data indicating that these protein modules maintain the two chains in an open conformation (54, 55). A similar effect has been proposed for the non-catalytic neogaroooligosaccharide binding site in the catalytic module of the GH16 β-agarase A from Z. galactanivorans, which is in a parallel orientation to the active site of the enzyme (25).

It is interesting to note that Aga86E is an exo-acting β-agarase, indicating that CBM6 of the enzyme binds to an agarose chain that is distinct from the sugar polymer hydrolyzed by the catalytic module. Indeed, it is tempting to speculate that by binding to the end of an agarose molecule the CBM6 helps to open up the double helical structure making the chain that is not appended to the CBM accessible to attack by the appended GH86 catalytic module. By contrast, the catalytic module in Aga16B is endo-acting and thus binds to internal regions of its target substrate. It is likely that the molecular dimensions of the modular agarase enable the catalytic module to access essentially internal regions of an agarose chain, such that all the available subsites in the substrate binding cleft are occupied with galactose or anhydro-galactose sugars. This combination of endo-acting catalytic modules, linked to CBMs that bind to the termini of polysaccharides, has been reported in two glycoside hydrolases: a GH10 xylanase that contains a CBM9, which binds to the reducing ends of both cellulose and xylan (56), and an endo-acting GH13 laminarase in which the CBM6 binds to the non-reducing end of β1,3-glucan chains (21). It is unclear why some endo-acting glycoside hydrolases are targeted by their cognate CBMs to the termini of polysaccharides. It is possible that these enzymes are directed to regions of the plant cell wall that have been subjected to prior β-agarase attack. Such areas of these recalcitrant composite structures may be highly disordered, and thus the CBMs are not just targeting glycoside hydrolases to their target substrates but regions of these molecules that are particularly amenable to enzyme attack.

Note Added in Proof—The biochemical properties of Aga86E and Aga16B are now described in Ref. 61.
