Alpha-agarases define a new family of glycoside hydrolases, distinct from beta-agarase families
Didier Flament, Tristan Barbeyron, Murielle Jam, Philippe Potin, Mirjam Czjzek, Bernard Kloareg, Gurvan Michel

To cite this version:
Didier Flament, Tristan Barbeyron, Murielle Jam, Philippe Potin, Mirjam Czjzek, et al.. Alpha-agarases define a new family of glycoside hydrolases, distinct from beta-agarase families. Applied and Environmental Microbiology, American Society for Microbiology, 2007, 73 (14), pp.4691-4694. <10.1128/AEM.00496-07>. <hal-00616725>

HAL Id: hal-00616725
http://hal.univ-brest.fr/hal-00616725
Submitted on 5 Sep 2011
Short communication

Alpha-agarases define a new family of glycoside hydrolases, distinct from beta-agarase families

Didier FLAMENT§, Tristan BARBEYRON, Murielle JAM, Philippe POTIN, Mirjam CZJZEK, Bernard KLOAREG and Gurvan MICHEL*

Centre National de la Recherche Scientifique; Université Pierre et Marie Curie-Paris6, Unité Mixte de Recherche 7139 "Marine Plants and Biomolecules", Station Biologique, F-29682 Roscoff Cedex, Bretagne, France

Keywords: marine bacteria, agar, glycoside hydrolase, CBM6, calcium-binding module

Running title: Alpha-agarases define a new family of glycoside hydrolases

*Corresponding author: Tel (33) 298 29 23 30, Fax (33) 298 29 23 24, E-mail: gurvan@sb-roscoff.fr

§ Current address: UMR 6197 (IFREMER / CNRS), Laboratoire de Microbiologie et Environnements Extrêmes, Plouzane, Bretagne, France

Abbreviations: aa, amino acid(s); bp, base pair(s); kb, kilobase(s); nt, nucleotide, kDa, kiloDalton; CBM(s), carbohydrate-binding module(s); GH(s), glycoside hydrolase(s); SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis, ORF, open reading frame; TSP3, thrombospondin type 3 repeats.
ABSTRACT

The gene of the $\alpha$-agarase from “Alteromonas agarilytica” has been cloned and sequenced. The gene-product (154 kDa) is unrelated to $\beta$-agarases and instead belongs to a new family of glycoside hydrolases (GH96). The $\alpha$-agarase also displays a complex modularity, with the presence of five TSP3 repeats and three CBM6s.
Agars are the main cell wall components of numerous red macroalgae. These polymers consist of 3,6-anhydro-L-galactoses and D-galactoses alternatively linked by α-(1,3) and β-(1,4) linkages (10). Agars constitute a crucial carbon source for a number of marine bacteria which secrete agarolytic enzymes, mainly β-agarases (EC 3.2.1.81) which hydrolyze the β-(1,4) linkages (See for a review, 12). They are found in three distinct families of glycoside hydrolases, families GH16, GH50 and GH86 (CAZY database, 6). Structural data are only available for the GH16 β-agarases ZgAgaA and ZgAgaB from Zobellia galactanivorans (1, 2, 9). In contrast, “Alteromonas agarilytica” secretes an α-agarase (EC 3.2.1.158) which randomly hydrolyzes the α-(1,3) linkages in agars, releasing agarotetraose as its main end product (13, 15, 18). In this context, we sought to determine whether the agarases of the α-type share any structural relationships with β-agarases. We report here the cloning of the α-agarase gene agaA from “A. agarilytica” and we demonstrate that α-agarases define a new GH family.

**Cloning of the agaA gene.**

The α-agarase from “A. agarilytica” (AaAgaA) was purified as previously described (13) and N-terminal and internal peptide sequences were determined (Pasteur Institute, France): ETLELQAESFANSGG (A) and QPRVYNPNEHIVAEIQGPAT (B), respectively. With degenerated oligonucleotides derived from these microsequences (CARGCIGARTCITTYGCIAA and TAYAYCCNAAYGARCAAYAT, respectively), a 2.5 kb DNA fragment was amplified by PCR using A. agarilytica genomic DNA and labeled with [α-32P]dCTP. This radiolabeled probe was used to screen an “A. agarilytica” genomic library, prepared as previously described (5). Among the ~5000 recombinant clones, two positive clones (pAA1 and pAA2) were identified with inserts of 7.4 kb and 17.9 kb in length. Southern blot analysis and plasmid mapping indicated that both inserts encompassed the same gene which is present in only one copy in the genome. Plasmid pAA1 was sequenced on both strands over 4651 bp and a single ORF
was identified and referred to as *agaA* (4287 bp). Potential -35 and -10 promoter regions (TTGAtc and TAcAca) and a Shine Dalgarno sequence (GGAG) were identified upstream of the start codon. A possible transcription termination codon was found downstream of the TAA stop codon. The deduced gene-product is a preprotein of 1429 residues (154 kDa) which includes the peptides A and B. A signal peptide cleaved between A26 and E27 is predicted by SIGNALP (4), consistent with the N-terminal sequencing of the purified extracellular α-agarase.

Several attempts were made to overexpress *AaAgaA* and its isolated modules in *E. coli* with pET or pGEX vectors in various conditions. Unfortunately, the constructs always yielded inclusion bodies. However, within one week of culture at 22°C on Zd agar broth (3), the *E. coli* clones harboring the plasmid pAA1 dug a hole in the substratum, indicating agar degradation. Therefore, under the control of its own promoter, *agaA* was successfully translated into an active, recombinant enzyme, confirming that this gene indeed encodes the α-agarase.

*AaAgaA* is a complex, modular protein with a catalytic domain defining a new GH family

Only the N-terminal region of *AaAgaA* displays significant sequence similarity with proteins in the UniProt database. Based on InterProScan (14), eight distinct modules were identified in this region (Fig. 1A), five thrombospondin type 3 repeats (TSP3-1, D171-G203; TSP3-2, D360-L392; TSP3-3, D393-A425; TSP3-4, D426-L458; TSP3-5, D459-G491) and three carbohydrate-binding modules from the family 6 (CBM6-1, E27-R159; CBM6-2, E209-T343; CBM6-3, S659-L792).

The closest characterized protein matching the TSP3 repeats of *AaAgaA* is the cellulase CelG from *Pseudoalteromonas haloplanktis*. In CelG, the TSP3 repeats constitute an extended linker connecting the GH5 catalytic module and a C-terminal CBM5 (17). The TSP3
repeats of *AaAgaA* present about 50% sequence identity with their counterparts in *CelG*. They also display ~30% sequence identity with the “true” type 3 repeats found in human thrombospondin, whose crystal structure has been solved (11). These modules lack secondary structures and are organized around a core of calcium ions coordinated by conserved aspartates (DxDxDGxx[D/N]xxDxC motif). The conserved cysteine is involved in a disulphide bridge linking adjacent TSP3 repeats, strengthening their stability (11). This motif is strictly conserved in each of the TSP3 repeats of *AaAgaA* (Fig. 1B), indicating that these modules adopt similar structure and likely bind calcium ions. This is consistent with the observations that α-agarase activity is stabilized by the presence of calcium ions (13, 18).

In BLASTp searches with the three CBM6s from *AaAgaA*, the highest E-values are always obtained for the CBM6 sequences tethered to β-agarases, while they significantly decrease with CBM6 linked to non-agarolytic enzymes. Only the CBM6s attached to the β-agarases *SdAga16B* and *SdAga86E* from *Saccharophagus degradans* were shown to actually bind agarose (8). A pairwise comparison indicates a strong sequence identity (51%) of CBM6-1 with the CBM6 from *SdAga16B* (Fig. 1C), while CBM6-2 and CBM6-3 are more divergent (28% and 26%, respectively). The crystal structure of the CBM6 from *SdAga16B* revealed that five residues are critical for the recognition of the non-reducing end of the agarose chain: Asn39, Tyr40, Trp97, Trp127 and Asn130 (8). Four of these residues are strictly conserved in CBM6-1, while Tyr40 is substituted by a similar aromatic amino acid, Phe64 (Fig. 1C). Altogether these results strongly suggests that CBM6-1 is an agar-binding module and likely displays selectivity towards the non-reducing termini of agarose chains. In contrast, the five critical residues are only partially conserved in CBM6-2 and CBM6-3. Therefore the specificity of these latter CBM6 is less certain.

Finally, a BLASTp search on the patent databank at the NCBI identified two proteins from a marine bacterium with a strong sequence identity with the C-terminal region of
AaAgA (49% and 77% respectively, Fig. 2). These proteins, also described as α-agarases (16), encompass three modules, two N-terminal CBM6s and the C-terminal module conserved with AaAgA. Since the N-terminal region of AaAgA encompasses only additional, non-catalytic modules, its conserved C-terminal region (Asn809-His1429) likely contains its active site. Together these three catalytic modules (~620 residues) constitute a new family of glycoside hydrolases, referred to as family GH96 (CAZY database). Therefore α-agarases are structurally unrelated to the β-agarases from the families GH16, GH50 and GH86.

ACCESSION NUMBERS

The nucleotide sequence of the α-agarase from “Alteromonas agarilytica” has been deposited in Genbank with the accession number AF121273. Its amino acid sequence is available in Swiss-Prot under the accession number Q9LAP7.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the help of the Conseil Régional de Bretagne, which supported this research through a Ph.D. fellowship to D. F. and a research grant (PRIR n° 98C389).
REFERENCES


FIGURE LEGENDS

Figure 1: A. Modular architecture of the α-agarase AaAgaA. CBM6 and GH96 refer to carbohydrate-binding modules of the family 6 and glycoside hydrolase module of the family 96, respectively. The grey boxes correspond to the thrombospondin type 3 repeats (TSP3). B. Structure-based alignment of the TSP3 repeats of the human thrombospondin (PDB code 1UX6), of the α-agarase AaAgaA and of the cellulase CelG from Pseudoalteromonas haloplanktis (trEMBL code: O86099). C. Structure-based alignment of the three CBM6s from the α-agarase AaAgaA and of the agar-specific CBM6s tethered to the β-agarases SdAga16B (Genpept: ABD80437) and SdAga86E (Genpept: ABD81915) from Saccharophagus degradans. These modules are compared to the secondary structures of the CBM6 appended to SdAga16B (PDB code 2CDO). Alpha helices and beta strands are represented as helices and arrows, respectively, and beta turns are marked with TT. The black triangles mark the residues involved in the recognition of the non-reducing end of agarose chains. Figure 1B, C and Figure 2 were prepared using the program ESPrit (7) and use the same colour codes.

Figure 2: Sequence alignment of the catalytic module of AgaA from A. agarilytica and of Aga14 and Aga15 (Patent US 6599729). Dark shaded boxes enclose conserved positions.