
Tristan Barbeyron,* Anita Gerard,* Philippe Potin,* Bernard Henri ssat,† and Bernard Kloareg*

*Centre d’Études d’Océanologie et de Biologie Marine, CNRS, Roscoff, France; and †Centre de Recherches sur les Macromolécules Végétales, CNRS, Grenoble, France

We report here cloning from the marine gliding bacterium *Cytophaga drobachiensis* of κ-carrageenase, a glycoside hydrolase involved in the degradation of κ-carrageenan. Structural features in the nucleotide sequence are pointed out, including the presence of an octameric Ω sequence similar to the ribosome-binding sites of various eukaryotes and prokaryotes. The cgkA gene codes for a protein of 545 aa, with a signal peptide of 35 aa and a 229-aa-long posttranslationally processed C-terminal domain. The enzyme displays the overall folding and catalytic domain characteristics of family 16 of glycoside hydrolases, which comprises other β-1,4-α-1,3-D-galactan hydrolases, β-1,3-D-glucan hydrolases (laminarinases), β-1,4-1,3-D-glucan hydrolases (licheninases), and β-1,4-D-xyloglucan endotransglycosylases. In order to address the origin and evolution of CgkA, a comprehensive phylogenetic tree of family 16 was built using parsimony analysis. Family-16 glycoside hydrolases cluster according to their substrate specificity, regardless of their phylogenetic distribution over eubacteria and eukaryotes. Such a topology suggests that the general homology between laminarinases, agarases, κ-carrageenases, licheninases, and xyloglucan endotransglycosylases has arisen through gene duplication, likely from an ancestral protein with laminarinase activity.

Introduction

The extracellular matrix polysaccharides of marine red algae, carrageenans and agaroses, consist of a glycosidic backbone of β-D-galactose residues linked by alternating (1→3) and (β1→4) linkages. In agaroses, the (1,4)-linked galactose units are in the L conformation, whereas they are in the D configuration in carrageenans (Rees 1969). A further layer of complexity is added by the number and the position of the sulfate substituents per disaccharide repeat unit and the occurrence of a 3,6-anhydro bridge in the (1,4)-linked galactose residue (fig. 1).

The structural gene of the κ-carrageenase from *Al teromonas carrageenovora* displays sequence similarity with members of family 16 of glycoside hydrolases (*Barbeyron, Henrissat, and Kloareg 1994*), which comprises a variety of β-1,3-glucanases, β-1,4-D-transglycosylases, and β-1,3,1,4-β-glucanases (Henrissat 1991; Henrissat and Bairoch 1993, 1996). These hydrolases display a double-displacement lytic mechanism, resulting in a net retention of the anomeric configuration of the reaction products and conferring the possibility of transglycosylation as a side reaction (Sinnott 1990; Potin et al. 1995). *Cytophaga drobachiensis*, a marine aerobic gram-negative gliding bacterium taxonomically distant from the genera *Alteromonas* and *Streptomyces*, produces κ-carrageenase (Potin et al. 1991). We report here the cloning, functional screening, and sequencing of the gene for this protein and show that, based on its structural and catalytic characteristics, it is a novel family-16 member. Its secondary-structure elements are compared with those of the κ-carrageenase from *A. carrageenovora* (Barbeyron, Henrissat, and Kloareg 1994) and with representatives of family-16 glycoside hydrolases, using the hydrophobic cluster analysis (HCA) procedure (Gaboriaud et al. 1987; Lemesle-Varloot et al. 1990). In addition, a phylogenetic analysis of a representative variety of family-16 members is presented, and we discuss the molecular evolution of these glycan hydrolases in relation to their substrate specificity.

Materials and Methods

Bacterial Strains and Culture Conditions

*Cytophaga drobachiensis* was isolated in our laboratory (Potin et al. 1991) from the red alga *Delesseria sanguinea* (Ceramiales, Rhodophyta). *Escherichia coli* strain DH5α (recA, endA1, gyrA96, thi1, hsdR17 [rk-mk+], supE44, relA1, lacZ ΔM15) was grown on Luria Bertani (LB) (Maniatis et al. 1982) at 37°C or 22°C on Zd medium (Bacto tryptone 5 g/l; yeast extract 1 g/l; NaCl 10 g/l, pH 7.2) supplemented with 1.5% κ-carrageenan (Mero-Rousselot-Satia). Antibiotics were added to solid or liquid media at 50 μg/ml ampicillin and 15 μg/ml tetracycline.

Cloning and Sequencing of the κ-Carrageenase

The internal peptide sequence of the κ-carrageenase, purified from *C. drobachiensis* according to Potin et al. (1991), was determined by the automatic Edman degradation method. Total DNA from *C. drobachiensis* was prepared as previously described (Barbeyron, Kean, and Forterre 1984), cut by the restriction endonuclease *Nde* II, and fractionated on a sucrose gradient. DNA fragments of approximately 4–10 kb were cloned at the BamHI site of plasmid vector pAT153, and the genomic library was screened for the presence of re-
combinant κ-carrageenase by plating the library on Zd medium solidified with κ-carrageenan.

Plasmid DNA was isolated from the clones showing κ-carrageenase activity by the alkaline lysis method (Birnboim and Doly 1979). The sizes and physical maps of inserts were determined by agarose gel electrophoresis after single and double digestion with various restriction enzymes. DNA fragments were extracted from agarose by the glass wool method (Heery, Gannon, and Powell 1990) and subcloned in phagemid pBluescript and are referred to as pKC1-κcarrageenan broth, five independent colonies had made a hole in the substratum. They are referred to as pKC1–pKC5. Inserts ranged from 5.2 to 11.1 kb and shared a common BamHI-Xba I fragment of 4.5 kb. Internal EcoRI and Nde I fragments (fig. 2), 1,100 and 600 bp long, respectively, were subcloned from pKC1 in pBluescript and are referred to as pKCE11 and pKCN6. They did not display a κ-carrageenase+ phenotype.

**Results and Discussion**

The κ-carrageenase gene of *Cytophaga drobachiensis*

The genomic library of *C. drobachiensis* DNA contained approximately 6,000 recombinant clones. Within 2 months of culture at 22°C on Zd κ-carrageenan broth, five independent colonies had made a hole in the substratum. They are referred to as pKC1–pKC5. Inserts ranged from 5.2 to 11.1 kb and shared a common BamHI-Xba I fragment of 4.5 kb. Internal EcoRI and Nde I fragments (fig. 2), 1,100 and 600 bp long, respectively, were subcloned from pKC1 in pBluescript and are referred to as pKCE11 and pKCN6. They did not display a κ-carrageenase+ phenotype.

Plasmids pKC1, pKCE11, and pKCN6 were used to determine, on both strands, the nucleotide sequence of the structural gene of κ-carrageenase. The pKC1 insert was sequenced over 4,425 bp. Its G+C content, 41.9 mol%, is consistent with the G+C ratio in the genus *Cytophaga* (30–45 mol%; Reichenbach 1989). It contained a single ORF of 1,635 bp, referred to as cgkA (fig. 2). In the 3' region, a stem-loop is found downstream of the TAA stop codon, followed by five thymidine residues. Its free energy (~118 kJ mol⁻¹) and the presence of the thymidine residues suggest that this sequence functions as a site for rho-independent transcriptional termination.

The deduced amino acid sequence of the κ-carrageenase from *C. drobachiensis* was aligned with that of *A. carrageenovora*, using HCA (Gaboriaud et al. 1987; Lemesle-Varloot et al. 1990). Secondary structural similarities were assessed by pairwise calculation of HCA scores. The rules described by Lemesle-Varloot et al. (1990) were used to discriminate false similarities from true similarities. Family-16 glycosidase hydrolases were aligned using Pileup (Wisconsin Package, version 8, Genetics Computer Group, Madison, Wis.), with a gap creation penalty of 3.0 and a gap extension penalty of 0.1, and the alignment was manually refined according to the results of HCA. A parsimony tree was then built using the PROTPARS program of PHYLIP 3.5c, and its reliability was tested by bootstrap analysis (SEQBOOT program of PHYLIP 3.5c; Felsenstein 1993) using 100 resamplings of the data set.

**Accession Number**

The nucleotide sequence encoding the κ-carrageenase of *C. drobachiensis* has been deposited in the GenBank database under the accession number AF007559.

**Plasmids pKC1, pKCE11, and pKCN6 were used to determine, on both strands, the nucleotide sequence of the structural gene of κ-carrageenase.**

The pKC1 insert was sequenced over 4,425 bp. Its G+C content, 41.9 mol%, is consistent with the G+C ratio in the genus *Cytophaga* (30–45 mol%; Reichenbach 1989). It contained a single ORF of 1,635 bp, referred to as cgkA (fig. 2). In the 3' region, a stem-loop is found downstream of the TAA stop codon, followed by five thymidine residues. Its free energy (~118 kJ mol⁻¹) and the presence of the thymidine residues suggest that this sequence functions as a site for rho-independent transcriptional termination.

Two hexamers, TATAAT and gTGACA, consistent with *E. coli* ‘‘−35’’ and ‘‘−10’’ consensus promoters and separated by 15 nucleotides, are found 257 nt upstream of the putative start codon. However, no canonical Shine-Dalgarno sequence can be identified in the ribosomal binding site region. In contrast, an octamer, AACATTA, is present 12 bp upstream of the ATG codon. This sequence, referred to as omega (Ω), is identical to the untranslated 5' leader region of tobacco mosaic virus RNA, which has been described as a universal enhancer of translation in both eukaryotic and prokaryotic organisms (Sleat et al. 1987; Gallie and Kado 1989; Firpo and Dahlberg 1990; Ivanov et al. 1992). The hypothesis that this motif may serve as the ribosome-binding site in *C. drobachiensis* is further supported by the repeat of the motif AACAA upstream of the Ω sequence. The Ω sequence does not contain any guanidine residue and, therefore, cannot interact with the 3' end of the 16S rRNA of *E. coli*. It has been found, however, that the ACAATTAC motif complements a region at positions 1340–1360 in *E. coli* 16S rRNA (Ivanov et al. 1992, 1995).

The predicted product of the κ-carrageenase gene is a protein of 545 aa with a theoretical molecular weight of 61,466 Da. The deduced amino acid sequence includes the internal peptide determined from...
**Fig. 2.** Nucleotide and deduced amino acid sequences of the \( \kappa \)-carrageenase gene from *Cytophaga drobachiensis* (cgkA). Restriction sites are indicated above the nucleotide sequence. Putative `-35` and `-10` boxes in the promoter region and the potential ribosome-binding site referred to as `Ω` are indicated in bold. The coding region starts at nucleotide 875. The proposed 35-aa signal peptide is shown in italics. The underlined amino acids were confirmed by protein sequencing.
The microsequencing of the purified κ-carrageenase (underlined in Fig. 2). As shown by hydropathy analysis, the N-terminus of the protein stands out as a domain with high hydrophobicity, suggesting that this domain is the signal peptide (Von Heijne 1983, 1985). In accordance with the “(−3, −1) rule” (Von Heijne 1983), the cleavage site of the signal peptide is assigned between Thr-35 and Ser-36, with the initiation codon at position 875. Such an exceptional length for a signal peptide may be accounted for by the length of its own N-terminal hydrophilic segment (13 aa). The presence of an arginine residue at position 12, two amino acids downstream of the second methionine residue, allows for the possibility that ATG at position 902 is the initiation codon. However, two Lys immediately follow the putative first methionine residue, a classical feature in signal peptides, and we propose that the signal peptide ends at Thr-35.

The molecular weight of the protein calculated after removal of the signal peptide, ca. 57,400 Da, is higher than the molecular weight initially determined by SDS-PAGE electrophoresis for the purified extracellular κ-carrageenase of *C. drobachiensis* (Thr-316, this work) and that of *A. carrageenovora* (Asn-301, Potin et al. 1995) are underlined.
cleavage of the signal peptide (after Thr-35), the C-terminal end is probably processed off at Thr-316. Interestingly, the C. drobachiensis and A. carrageenovora κ-carrageenase sequences strongly diverge after that position (fig. 3).

Structural Relationships with the Other Family-16 Glycoside Hydrolases

The genus Cytophaga belongs to the Cytophaga/Flexibacter/Bacteroides group, whereas Alteromonas is a proteobacterium (Woese 1987). However, the κ-carrageenase from C. drobachiensis shares significant sequence similarities with that from A. carrageenovora, with an overall identity of 36% (fig. 3), showing that it is a novel member of family 16 of glycoside hydrolases. Three domains are quite similar in the primary sequences, from Tyr-158 to Leu-166, from Trp-316 to Pro-265, and from Pro-302 to Arg-302 (numbering of the C. drobachiensis κ-carrageenase sequence). In particular, within one of the best conserved motifs, two glutamic acid residues are present and separated by 4 amino acids (Glu-160 and Glu-165 in the C. drobachiensis sequence). This organization is characteristic of the catalytic site of family-16 glycoside hydrolases, which features two glutamic acid residues separated either by three amino acids, in the bacterial lichenases and plant xyloglucan endotransglycosylases, or by four residues, in the other galactanases and laminarinases. The catalytic function of the equivalent Glu-134 and Glu-138 in Bacillus licheniformis β-1,3,1,4-glucanase (Planas et al. 1992; Juncosa et al. 1994) and of Glu-105 in B. amyloliquificaciens β-1,3,1,4-glucanase (Hojl et al. 1992) was demonstrated by site-directed mutagenesis and selective inhibition. We therefore assume that the equivalent Glu-160 and Glu-165 in C. drobachiensis κ-carrageenase constitute its catalytic machinery, as the nucleophile and the acid/base residues, respectively.

The HCA plots of the κ-carrageenase from C. drobachiensis and of other representatives of family-16 glycoside hydrolases, namely the β-agarase (α-1,3-L-β-1,4-d-galactanase) from Streptomyces coelicolor (Buttner, Fearnley, and Bibb 1987), the κ-carrageenase (α-1,3-d-β-1,4-d-galactanase) from A. carrageenovora (Barbeyron, Henrissat, and Kloareg 1994), the laminarinase (β-1,3-d-glucanase) from Rhodothermus marinus (Spillaert et al. 1994), the lichenase (β-1,3,1,4-d-glucanase) from Bacillus subtilis (Murphy, McConell, and Cantwell 1984), and the xyloglucan endotransglycosylase (β-1,4-d-transglycosylase) from Arabidopsis thaliana (Okazawa et al. 1993), are shown in figure 4. The two catalytic Glu residues were taken as the anchor points for the HCA comparison, and clusters were delineated based on the known three-dimensional structure of the Bacillus macerans lichenase (Keitel et al. 1993). HCA plots reveal a similar distribution of the hydrophobic clusters over a length of more than 150 residues, confirming that these enzymes share similar secondary-structure elements. Although identity scores calculated from the resulting pairwise alignments were all below 26% between glycosidases with different substrate specificities, the HCA scores based on the correspondences shown in figure 4 were above 60% (table 1), a significant similarity threshold in this comparison procedure (Lemesle-Varloot et al. 1990).

Thirteen distinct structural segments (I–XIII) are apparent in the HCA plots (fig. 4), among which I, II, IX, X, XI, and XIII appear to be best conserved. The two κ-carrageenases and the β-agarase from S. coelicolor share a common segment (VII) which is not present in the other family-16 members. Segment VI is found only in the β-agarase from S. coelicolor (fig. 4) as well as in other agarases of family 16 (unpublished data). The hydrophobic segment (III) around the catalytic residues appears to be well conserved among galactanases (agarase and κ-carrageenases) and, to a lesser extent, among laminarinases.

Table 1

<table>
<thead>
<tr>
<th>Sequences</th>
<th>KapAc</th>
<th>KapCd</th>
<th>AgarSc</th>
<th>LamRm</th>
<th>LichBs</th>
<th>XetAt</th>
</tr>
</thead>
<tbody>
<tr>
<td>KapAc</td>
<td>100</td>
<td>71</td>
<td>71</td>
<td>71</td>
<td>72</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>45</td>
<td>24</td>
<td>19</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>KapCd</td>
<td>100</td>
<td>62</td>
<td>67</td>
<td>67</td>
<td>61</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>17</td>
<td>21</td>
<td>18</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>AgarSc</td>
<td>100</td>
<td>61</td>
<td>65</td>
<td>64</td>
<td>61</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>20</td>
<td>14</td>
<td>18</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>LamRm</td>
<td>100</td>
<td>65</td>
<td>71</td>
<td>68</td>
<td>61</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>12</td>
<td>13</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>LichBs</td>
<td>100</td>
<td>71</td>
<td>61</td>
<td>61</td>
<td>61</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>XetAt</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

a For each entry, the average HCA score is given above, and the sequence identity score is given below in italics. Pairwise HCA scores (Gaboriaud et al. 1987; Lemesle-Varloot et al. 1990) were calculated for each cluster as: HCA score = 2C0/(C1 + C2) × 100%, where C1 and C2 are the numbers of hydrophobic residues in sequences 1 and 2, respectively, and CR is the number of hydrophobic residues in sequence 1 that are in correspondence with those in sequence 2. The final HCA score was the average score of all the hydrophobic clusters along the 13 structural segments delineated in figure 4.

b Abbreviations: KapAc, κ-carrageenase of Alteromonas carrageenovor; KapCd, κ-carrageenase of Cytophaga drobachiensis; AgarSc, β-agarase of Streptomyces coelicolor; LamRm, laminarinase of Rhodothermus marinus; LichBs, lichenase of Bacillus subtilis; XetAt, xyloglucan endotransglycosylase of Arabidopsis thaliana.
Fig. 5.—Parsimony phylogenetic tree of the family-16 glycoside hydrolases, including various xyloglucan endotransglycosylases, laminarinas, lichenases, and galactanases. Numbers indicate the bootstrap values in the parsimony analysis. The distance matrix tree (not shown) was similar to the parsimony tree. Abbreviations: LamRm, laminarinase of *Rhodothermus marinus* (SWISS-PROT U04836); LamBcglA, *glcA* laminarinase gene of *Bacillus circulans* (SWISS-PROT P23903); LamBcglh, *bglH* laminarinase gene of *B. circulans* (DDBJ D17519); LamTn, laminarinase of *Thermotoga neapolitana* (EMBL Z47974); LamCt, laminarinase of *Clostridium thermocellum* (EMBL X89732); LamV1Ch, laminarinase of *Paramecium bursaria Chlorella virus1* (NCBI 113143); LamSp, laminarinase of *Oerskovia xanthineolytica* (GenBank U49711); LamOx, laminarinase of *Oerskovia xanthineolytica* (GenBank U56935); LamPr, laminarinase of *Pyrococcus furiosus* (Gueguen et al., 1997); FGaTt, horseshoe crab clotting factor G subunit alfa of *Tachypleus tridentatus* (SWISS-PROT D16622); KRE6Sc, β-glucan synthesis-associated protein of *Saccharomyces cerevisiae* (NCBI 113143); SKN1Sc, β-glucan synthesis-associated protein of *Saccharomyces cerevisiae* (SWISS-PROT P33336); LichBm, lichenase of *Bacillus macerans* (SWISS-PROT P23903); LichBp, lichenase of *Bacillus polymyxa* (SWISS-PROT P04957); LichBl, lichenase of *Bacillus licheniformis* (SWISS-PROT P27051); LichBb, lichenase of *Bacillus brevis* (SWISS-PROT P37073); LichSb, lichenase of *Streptococcus bovis* (EMBL Z92911); LichBsp, lichenase of *Bacillus sp.* (NCBI 113143); LichCl, lichenase of *C. thermocellum* (SWISS-PROT X63355); βGluCt, β-glucanase of *C. thermocellum* (SWISS-PROT P29716); LichRf, lichenase of *Ruminococcus flavefaciens* (PIR A36910); LichFs, lichenase of *Fibrobacter succinogenes* (SWISS-PROT P17989); Meri5At, Arabidopsis thaliana protein (SWISS-PROT P24806); BRU1Gm, brassinosteroid-regulated protein of *Glycine max* (GenBank L22162); XETB1Le, xyloglucan endotransglycosylase B1 (XET B1) from fruits of *Lycopersicon esculentum* (GenBank X82685); XETB2Le, XET B2 from fruits of *L. esculentum* (GenBank X82684); XETLe, XET from shoots of *L. esculentum* (GenBank D16456); XETZm, XET of *Zea mays* (GenBank U15781); XETTa, XET of *Triticum aestivum* (GenBank D16457); XETGh, XET of *Gossypium hirsutum* (GenBank D88413); XETGm, XET of *G. max* (GenBank D16455); XETVa, XET of *Vigna angularis* (GenBank D16458); XETAt, XET of *A. thaliana* (GenBank D16454); XETNi, XET of *Nicotiana tabacum* (DDBJ D86730); XETHv, XET of *Hordeum vulgare* (EMBL X91659); XEAHv, XET-like protein of *Hordeum vulgare* (EMBL X91747); XEBHv, XET-like protein of *H. vulgare* (EMBL X9175); PM2Hv, XET-like protein of *H. vulgare* (EMBL X91660); PM3Hv, XET-like protein of *H. vulgare* (EMBL X9175).
Phylogenetic Analysis of Family-16 Glycoside Hydrolases

Among the 57 families of glycoside hydrolases identified so far on the basis of amino acid sequence similarities (Henrissat and Bairoch 1996), family 16 is one of the few that encompass enzymes with widely different substrate specificities. In order to further analyze the relatedness of family-16 glycoside hydrolases, the amino acid sequences of 57 members of this family were aligned on the basis of the HCA results, yielding a range of 178–193 informative positions. Besides agarases and κ-carrageenases (EC 3.2.1.81 and EC 3.2.1.83), which cleave the β-1,4 linkages of β-1,4-α-1,3-galactans, the alignment includes laminaranases (EC 3.2.1.39), which hydrolyse β-1,3-glucosyl linkages within homo-β-1,3-α-glucan sequences; lichenases (EC 3.2.1.73), which split the β-1,4 linkages adjacent to β-1,3-linked glucosyl residues in the β-1,4-1,3-α-glucans known as lichenans (Parrish, Perlin, and Reese 1966); and various xyloglucan endotransglycosylases, the amino acid sequences of 57 members of this family were aligned on the basis of the HCA results, yielding a range of 178–193 informative positions. Besides agarases and κ-carrageenases, the amino acid sequences of 57 members of this family were aligned on the basis of the HCA results, yielding a range of 178–193 informative positions. 

In the unrooted parsimony tree inferred from this alignment (fig. 5), family-16 glycoside hydrolases appear as four distinct lineages, coincident with their substrate specificity. This topology suggests that both the transfer of segments of β-1,4-xyloglucan molecules to other xyloglucan molecules in plant cell walls (Nishitani and Tominaga 1992; Rose, Brummel, and Benett 1996).

Acknowledgments

This is a contribution of GDR 1002, “Biology, Biochemistry and Genetics of Marine Macroalgae.” The authors are grateful to Yves Pétillot (Institut de Biologie Structurale, Grenoble) for mass-spectroscopy analysis of the κ-carrageenase of C. drobachiensis. We also thank Y. Gueguen for sharing with us unre-
leased sequence data on the laminarinase of *Pyrococcus furiosus*.

**LITERATURE CITED**


MITIKO GO, reviewing editor

Accepted February 10, 1998