Cah. Biol. Mar. (2001) 42 : 169-183

The sulphated-galactan hydrolases, agarases and carrageenases: structural biology and molecular evolution

Tristan BARBEYRON[1], Didier FLAMENT[1], Gurvan MICHEL[2], Philippe POTIN[1] and Bernard KLOAREG[1]
[1]*UMR 1931 CNRS-Goëmar, Station Biologique CNRS-INSU-Université Paris 6,*
Place Georges-Teissier, BP 74, F29682 Roscoff Cedex, France.
[2]*Laboratoire de Cristallographie Macromoléculaire, Institut de Biologie Structurale, CNRS/CEA,*
41 Avenue des Martyrs, 38027 Grenoble Cedex 1, France
**Corresponding author: Tristan Barbeyron, Fax 33 2 98 29 23 24; E-mail: barboun@sb-roscoff.fr.*

Abstract: The carrageenans and agars are major cell-wall polysaccharides from red algae. These sulphated galactans are degraded by enzymes, called carrageenases and agarases that display strict substrate specifities and recognize the pattern of galactan sulphation. From a set of various marine bacteria enzymes, we have investigated the influence of ester-sulphate groups, of D/L isomery and of linkage anomery on the structure-function relationships of the specific galactan hydrolases that degrade sulphated polysaccharides. With this aim, we have cloned a representative set of sulphated-galactan hydrolase genes. The sequence analysis methods indicate that the β-agarases and κ-carrageenases display secondary structure similarities with members of family 16 of glycoside hydrolases. In contrast, the ι-carrageenases have no structural relationships with the family-16 β-agarases and κ-carrageenases and they constitute a novel structural family of glycan hydrolases. As a preliminary step towards the functional analysis of these two structural families, we have overexpressed the ι- and κ-carrageenase genes in *Escherichia coli* and crystals from these enzymes have been obtained. Finally, an α-agarase, the only one galactanase known to cleave the α-1,3 linkage in agarose has no similarity with other glycoside hydrolases or proteins and display some interesting characteristics. To date, this enzyme is an unclassified glycoside hydrolase.

Résumé : Les carraghénanes et les agars sont les polysaccharides pariétaux majeurs chez les algues rouges. Ces phycocolloïdes sont des galactanes sulfatés, hydrolysés par des enzymes appelées carraghénases et agarases. Ces galactanases ont une grande spécificité de substrat et semblent reconnaître le profil de sulfatation de leur substrat. Nous avons étudié l'influence du nombre et de la position des radicaux sulfate ainsi que celle de l'énantiomérie et de l'anomérie sur les relations structure-fonction des galactanases bactériennes dégradant les polysaccharides sulfatés. Les analyses de séquences des gènes clonés de galactanases indiquent que les β-agarases et κ-carraghénases montrent des similitudes de structure avec des membres de la famille 16 des glycoside hydrolases. Au contraire, les ι-carraghénases n'ont aucune relation structurale avec les galactanases de la famille 16 et forment une nouvelle famille de glycanases. Afin de réaliser une analyse fonctionnelle des galactanases de ces deux familles, nous avons surexprimé une ι- et une κ-carraghénase dans *Escherichia coli* et des cristaux ont été obtenus. Enfin, nous avons étudié une α-agarase, la seule galactanases connue à ce jour, qui hydrolyse la liaison α-1,3 dans l'agarose. Cette α-agarase ne présente aucune similitude de structure avec d'autres protéines et montre des caractéristiques originales. Pour l'heure, elle est non classée.

Keywords: agarase, carrageenase, cell-wall, red algae, sequence alignment, glycoside hydrolases.

Introduction

Agars and carrageenans exhibit unique rheological properties and are widely used as texturing agents in various industries (De Ruiter & Rudolph, 1997). As reported by Lahaye (this issue), they consist of a linear backbone of galactose residues linked by alternating α-1,3 and β-1,4 linkages (Fig. 1). In agars the α(1,4)-linked galactose units are in the L-configuration whereas they are in the D-configuration in carrageenans (Rees, 1969). A further layer of complexity is the occurrence of a 3,6-anhydro bridge in the α(1,4)-linked galactose residue and the number (from 0 in agarose to 3 in λ-carrageenan) of sulphate substituents per digalactose repeating unit (Fig. 1).

The use of specific enzymes has been of great help in the elucidation of the fine structure of these galactans in combination with ^{13}C-NMR (Bellion et al., 1981; Lahaye & Rochas, 1991), as it provides a mean to enrich with the minor components that are below the detection limit of spectroscopic methods.

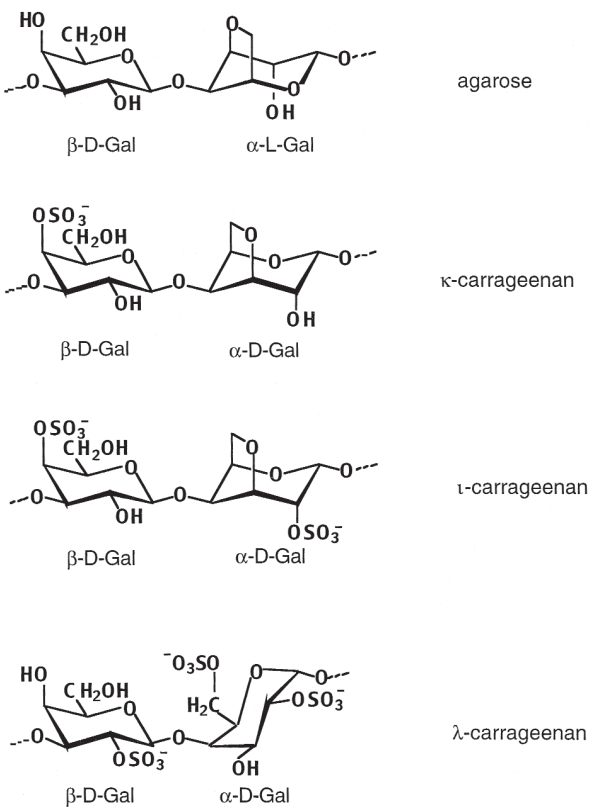


Figure 1. Disaccharide repeating units of agarose and carrageenans. Note that agarose is a neutral galactan, whereas κ-, ι- and λ-carrageenans are substituted by one, two, and three sulphate groups per repeating disaccharide, respectively.

Figure 1. Unités disaccharidiques répétitives de l'agarose et des carraghénanes. Observez que l'agarose est un galactane neutre, alors que κ-, ι- et λ-carraghénanes sont substitués par un, deux et trois radicaux sulfate par unité disaccharidique répétitive, respectivement.

Depending on their structure, sulphated galactans display anticoagulant (Carlucci et al., 1997), antiviral (Carlucci et al., 1997; Yamada et al., 1997), antitumoral (Hoffman, 1993), and immunomodulatory properties (Di Rosa, 1972; Yoshizawa et al., 1995). Carrageenan oligosaccharides also displayed defence-stimulating activities towards plant cells (Patier et al., 1995) and therefore oligogalactans may find valuable applications as active biocompounds in agriculture, cosmetics, or pharmaceuticals.

In contrast with chemical hydrolysis, enzymatic hydrolysis does not temper with the fine structure of the original substrates, and the potential of polysaccharidases for the production of novel intermediates of commercial interest is vast. With the ultimate goal of developing tools for the biotechnological production of sulphated oligogalactans, we have purified and cloned a variety of agarases and carrageenases. We review here the biochemical and molecular characteristics of these glycoside hydrolases and discuss their structural features in relation to their substrate specificity.

Biochemical features of galactan hydrolases

We maintain in our laboratory various marine bacteria that secrete galactan endo-hydrolases, including producers of κ-carrageenases, ι-carrageenases, and λ-carrageenases as well as of various agarases (Table 1). These bacteria all belong to the genera *Pseudoalteromonas* or *Alteromonas* and to the genus *Zobellia*, two groups of phylogenetically distant bacteria. All but one of these hydrolases are endo-β-galactanases. They cleave the internal β-1,4 linkages of agars or of carrageenans, yielding oligogalactans of the neocarrabiose or neoagarobiose series (Table 1). As these galactan hydrolases display a strict substrate specificity,

Table 1. Marine bacteria collection and their galactanase activities. The end products for each enzyme are indicated.

Tableau 1. Collection de bactéries marines produisant des activités galactanases. Le produit final d'hydrolyse pour chaque enzyme est indiqué.

Marine bacteria	Enzymes	End products
Pseudoalteromonas carrageenovora	κ-carrageenase	κ-neocarrabiose
Pseudoalteromonas atlantica	β-agarase I	Neoagarotetraose
	β-agarase II	Neoagarobiose
Alteromonas fortis	ι-carrageenase	ι-neocarratetraose
Alteromonas beaufortensis	β-agarase	Neoagarotetraose
Alteromonas agarlyticus	a-agarase	Agarotetraose
Zobellia galactanovorans	κ-carrageenase	κ-neocarratetraose
	ι-carrageenase	ι-neocarratetraose
	β-agarases A and B	Neoagarotetraose

they obviously recognize the pattern of sulphation on the digalactose repeating units. Galactan hydrolases were purified to electrophoretic homogeneity and their products were analysed by gel filtration followed by ^{13}C- and ^{1}H-NMR spectroscopy (Potin et al., 1991; Rochas et al., 1994). These polysaccharidases give rise to a variety of end products and can be used to generate a number of unique oligosaccharide series. In addition the κ-carrageenase from *Pseudoalteromona carrageenovora* was shown to display endotransglycosylating activity (Potin et al., 1995), a property that may be used to catalyse the transfer of various substituents onto oligo-κ-carrageenans.

Molecular evolution of galactan hydrolases

After purification of these galactanases, we have investigated their substrate specificity. It appears that they display a high substrate specificity. An agarase does not recognize carrageenan as substrate and within carrageenases, a κ-carrageenase does not cleave the ι-carrageenan (Potin, 1992; Potin et al., 1991). These observations suggest that the bacterial galactanases recognize the pattern of sulphation on the digalactose repeating unit of galactans. Sulphonic groups are bulky substituents which are likely to hinder access to the glycosidic linkages. This set of glycoside hydrolases therefore provides a unique opportunity to investigate the influence of ester-sulphate groups on the structure-function relationships of the hydrolases that degrade sulphated polysaccharides.

To study the molecular bases of the specificity of galactanases, we have cloned and sequenced the corresponding structural genes. The strategy used was to purify the genomic DNA from the marine bacteria, to cleave it partially, to clone the different fragments obtained and to plate the resulting genomic library onto gels made with the galactan substrate of the enzyme under investigation. This functional screening allowed us to detect the positive clones as those which dug a hole in the substratum (Barbeyron, 1993; Barbeyron et al., 1994).

The β-agarases proteins from Zobellia galactanovorans

We have cloned two different agarase genes *agaA and agaB* from *Zobellia galactanovorans*, a marine bacterium previously referred to as *Cytophaga drobachiensis* strain Dsij (Barbeyron et al., 2001)

β-agarase A features

The *agaA* gene encode for a protein of 539 amino acids with a theoretical molecular mass of 60.0 kDa. It is synthesized as a preproprotein with a 19 residues signal peptide which allow translocation in periplasmic space after cleavage by the signal peptidase I between Ala19 and Ala20. Then, a second processing remove a C-terminal fragment from ca. Asn295, probably to go across the outer membrane of the cell. Indeed, the protein AgaA purified from the cell-free supernatant display a molecular mass of 31 kDa (Flament et al., unpublished) and it is capable to hydrolyse the neoagarododecaose down to neoagarotetraose.

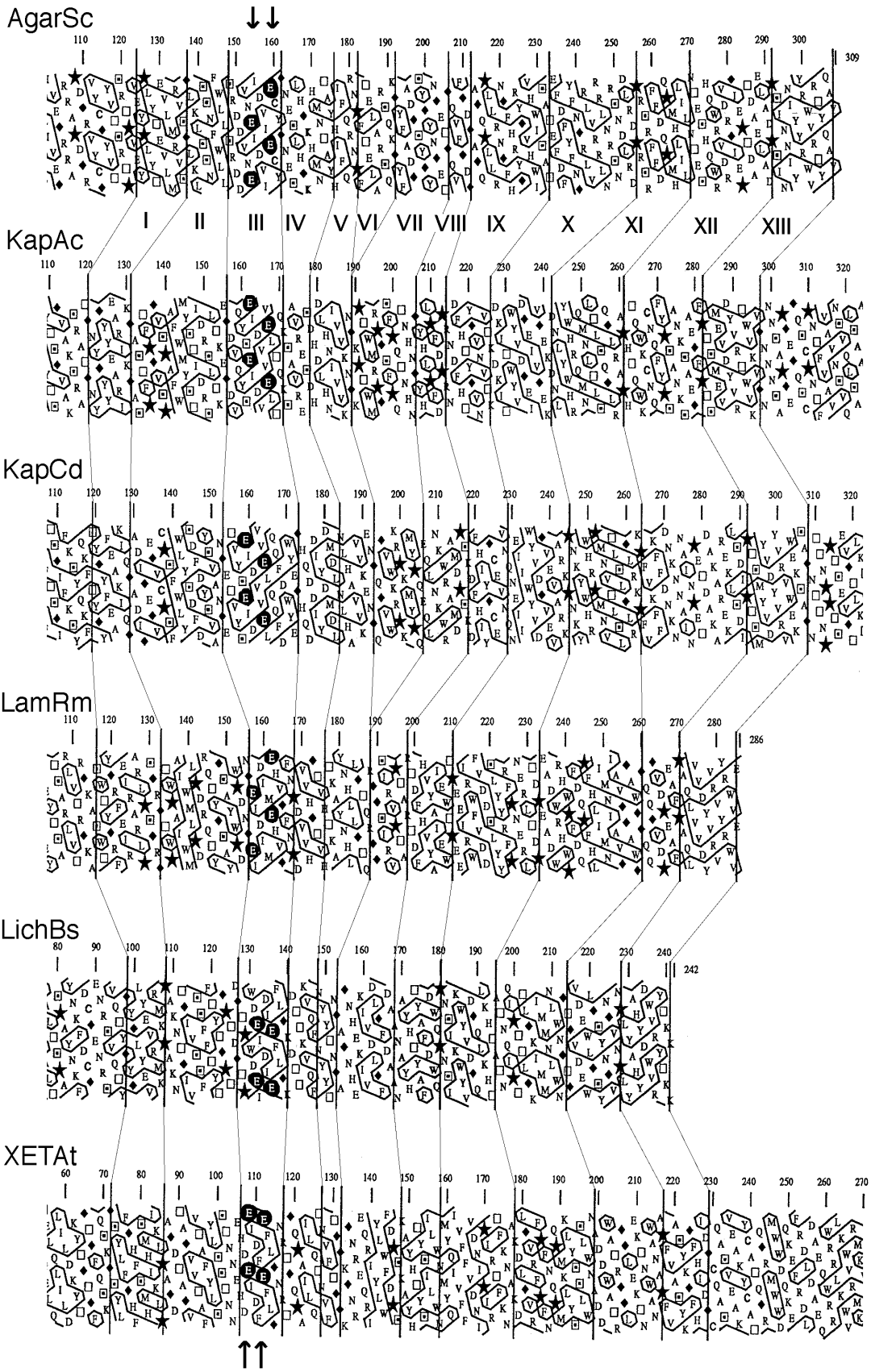
The β-agarase B features

The *agaB* gene encodes for a protein of 354 residues with a theoretical molecular mass of 40.68 kDa. The product of the *agaB* gene was not detected in the supernatant. Analysis of the signal peptide characteristics seem to indicate that AgaB is a plasma membrane-associated lipoprotein (Flament et al., unpublished), with a probable cleavage site by the signal peptidase II between Gly17 and Cys18.

It is likely that AgaB is a membrane-bound, periplasmic protein whereas AgaA is extracellular and binds to agarose. This agarolytic system is reminiscent of that from *Pseudoalteromonas atlantica*, in which the periplasmic neoagarotetraose hydrolase, also known as agarase II, hydrolyses the products of the extracellular agarase known as agarase I (Morrice et al., 1983a, b).

Sequence comparisons

The two β-agarases from *Zobellia galactanovorans* are very similar in their primary structures, with 44.5% and 67.5% of identity and similarity, respectively. Significant similarities are also observed between AgaA or AgaB with the β-agarase I from *Pseudoalteromonas atlantica* (GeneBank M73783) and with *Streptomyces coelicolor* agarase (Buttner et al., 1987). The homology between both proteins was further investigated by hydrophobic cluster analysis (HCA). This method, which was developed for the comparison of proteins that are weakly related in their primary structure, rests upon the duplicated representation of the amino acid sequences on an α-helix two-dimensional pattern. In this representation, the clusters of hydrophobic residues are thought of as representative of definite secondary structure elements and proteins that display a similar distribution of hydrophobic clusters with comparable sizes and shapes are considered as superimposable in their 3-D structures (Lemesle-Varloot, 1990). The HCA approach was used to delineate the structural homologies among glycoside hydrolases, which to date fall in more than 83 distinct structural families (Henrissat & Bairoch, 1996; a structure-based classification of glycoside hydolases is available on line at URL: http://afmb.cnrs-mrs.fr/~pedro/CAZY/.) Within each family, glycoside hydrolases display the same overall folding, similar active sites and identical catalytic mechanisms. The HCA analysis demonstrates that the β-agarases from *Zobellia galactanovorans* and β-agarases from *S. coelicolor* and *P. atlantica* belong to the structural family 16 of glycoside hydrolases (Fig. 2, where the


AgarSc
KapAc
KapCd
LamRm
LichBs
XETAt
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII

S. coelicolor β-agarase only is shown). In particular, the catalytic site features of this family are very well conserved in the primary structure of *Z. galactanovorans* agarases. By comparison with the lichenases (Planas et al., 1992; Bordier et al., 1992; Juncosa et al., 1994) and laminarinase (Krah et al., 1998), which are members of the family 16 of glycoside hydrolases, we can predict that the catalytic amino acids are Glu147 and Glu152 in AgaA and Glu184 and Glu189 in AgaB, acting as a proton donor and an acceptor, respectively (data not shown).

The κ-carrageenase proteins also belong to the family sixteen of the glycoside hydrolases

The κ-carrageenase from Pseudoalteromonas carrageenovora

Enzymatic degradation of κ-carrageenan to yield oligosaccharides of the neocarrabiose series is known since 1955 (Yaphe & Baxter, 1955). The responsible enzyme, the κ-carrageenase, has been purified from the cell-free supernatant of *Pseudoalteromonas carrageenovora* culture, and used in order to elucidate the structure of its carrageenan substrate (Weigl & Yaphe, 1966; McLean & Williamson, 1979, 1981; Bellion et al., 1981). The gene encoding for the κ-carrageenase has been cloned from a genomic library of this bacteria (Barbeyron et al., 1994). The deduced protein sequence is composed of 397 amino acids and display a theoretical molecular mass of 44. 21 kDa (Fig. 3). After removal of a signal peptide of 25 residues, the molecular mass shift down to 41.6 kDa which is larger than the molecular mass of the purified protein of 35 kDa as determined on SDS-PAGE, suggesting that the periplasmic protein is further processed when crossing outer membrane, as already seen for the β-

```
  1 MKKPNFYGKMGRTALSSLFYLFFLGLVYGQQPTKT↓SNPNDQWTIKWSASDEFN-KNDPDW   59
    ||          | | |            ||       |   |       ||||| | |
  1  MKPISIVAFPIPAISMLLLSAVSQA↓ASMQPPIAK--PGETWILQAKRSDEFNVK-DAT-   55

 60 AKWIK-TGNLPNTSAWKWNNQKNVKISNGIAELTM-RHNANNTPPDGGT---------Y   107
     ||    | |     | || |        |||    ||   |      |   ||            |
 56 -KWNFQTENYGVWS-WK-NENAT--VSNGKLKLTTKRESHQRTFWDGCNQQQVANYPLY   109

108 FTSGIFKSYQKFTYGYFEAKIQGADIGEGVCPSFWLYSDFDYSVA-NGETVYSEIDVVEL   166
     |||   ||      ||| || | ||     || | || ||   | |     |   |||||||||
110 YTSGVAKSRATGNYGYYEARIKGASTFPGVSPAFWMYSTIDRSLTKEGDVQYSEIDVVEL   169

167 QQFDWY-EGHQDDIYDMDLNLHAVVKENGQGVWKRPKMYPQEQLNKWR-AMDPSKDFHIY   224
     |      |   |    |    |   ||| ||   | ||    ||    |       ||  ||| |
170 TQKSAVRES--DH--DLH-NI--VVK-NGKPTWMRPGSFPQTNHNGYHLPFDPRNDFHTY   221

225 GCEVNQNEIIWYVDGVE-VARKPNKYWHRPMNVTLSLGLRKPFVKFFDNKNNAINPETDA   283
     |   |     | |||||  | |   |  | ||||  || ||| ||| |         | |    |
222 GVNVTKDKITWYVDG-EIVGEKDNLYWHRQMNLTLSQGLRAPHTQW---KCNQFYPSAN-   276

284 K-AREKLSDIPTSMYVDYVRVWEKSAGNTTNPPTSEVGTLKTKGSKLVIDHWDASTGTIS   342
     | | |      |||| |||||  | |    |    |                |        |
277 KSA-EGF---PTSMEVDYVRTWVKVGNNNSAPGEGQSCPNTFVAVNSVQLSAAKQTLRKG   332

343 AVSNNTKTGQYAGSVNNASIAQIVTLKANTSYKVSAFGKASSPGTSAYLGISKASNNELI   402
              |        |    |          |          ||    ||        |
333 QSTTLESTVLPNCATNKKVIYSSSNKNVATVNSAGVV-KAKNKGTATITVKTKNKGKIDKL 392

403 SNFEFKTTSYSKGEIEIRTGNVQESYRIWYWSSGQAYCDDFNLVEINSGASQLNENETET   462

393 TIAVN                                                          397


463 ALEKGIHIYPNPYKNGPLTIDFGKPFSGEVQITGLNGRTFLRRNVVDQTSVQLLESKSKF   522


523 KSGLYIVKISGPDGEVSKKILVE                                       545
```

Figure 3. Sequence alignment of the *Zobellia galactanovorans* (upper sequence) and the *Pseudoalteromonas carrageenovora* (lower sequence) κ-carrageenases. Arrow in sequences indicates the signal peptidase I cleavage sites. The aa sequence underlined was confirmed by microsequencing of the purified κ-carrageenase from *P. carrageenovora*. The putative catalytic glutamic acid residues are indicated in bold. The proposed sites for *C*-terminal processing of the κ-carrageenase of *Z. galactanovorans* (Thr316, Barbeyron et al 1998) and that of *P. carrageenovora* (Asn301, Potin et al. 1995) are underlined twice. Sequences show 36% of identity and 52% of similarities.

Figure 3. Alignement des séquences des κ-carraghénases de *Zobellia galactanovorans* (séquence supérieure) et de *Pseudoalteromonas carrageenovora* (séquence inférieure). La flèche insérée dans les séquences indiquent les sites de coupure de la signal peptidase I. Les acides aminés soulignés dans la séquence de *P. carrageenovora* ont été déterminés par microséquençage. Les acides glutamiques impliqués dans le processus de catalyse sont indiqués en gras. Les sites de coupure proposés, en C-terminal des séquences, pour les κ-carraghénase de *Z. galactanovorans* (Thr316, Barbeyron et al., 1998) et de *P. carrageenovora* (Asn301, Potin et al., 1995) sont soulignés deux fois. Les deux séquences ont 36 % d'identité et 52 % de similitude.

Figure 2. Hydrophobic cluster analysis (HCA) of family-16 glycoside hydrolases with different substrate specificities. The catalytic glutamic acid residues are shown in black circles and also indicated by vertical arrows at the top and the bottom of the figure. The various structural segments are delineated by vertical bars and identified by roman numerals.

Figure 2. Analyse des amas hydrophobes de différents membres de la famille 16 des glycoside hydrolases présentant différentes spécificité de substrat. Les acides aminés catalytiques sont montrés dans des cercles noirs et indiqués par des flèches en haut et en bas de la figure. Les différents segments structuraux sont délimités par des barres verticales et identifiés par des nombres romains.

AgarSc : β-agarase *Streptomyces coelicolor* ; KapAc : κ-carrageenase *Pseudoalteromonas carrageenovora*; KapCd : κ-carrageenase *Cytophaga drobachiensis* (*Zobellia galactanovorans*) ; LamRm : laminarinase *Rhodothermus marinus* ; LichBs : lichenase *Bacillus subtilis* ; XETAt : xyloglucan endotransglycosylase *Arabidopsis thaliana*.

agarase A from *Zobellia galactanovorans*. This result has been confirmed by purification of the non mature protein from a cellular extract of *P. carrageenovora* (Potin et al., 1995). Moreover, the precise mass determination by ESI/MS of the secreted protein (31.7 kDa) has indicated that the second processing is between Asn301 and Ser302. This enzyme hydrolyses the κ-carrageenan to the κ-neocarrabiose sulphate and the κ-neocarratetraose sulphate, as major products (McLean & Williamson, 1979; Rochas et al., 1983), with an overall retention of the configuration of the anomeric carbon. The κ-carrageenase also displays transglycosylating activity (Potin et al., 1995).

The κ-carrageenase from Zobellia galactanovorans

A κ-carrageenase activity has been purified also from a cell-free supernatant of *Zobellia galactanovorans* strain Dsij culture. The purified enzyme displays a molecular mass of 40 kDa (Potin et al., 1991) as determined by SDS-PAGE. It hydrolyses κ-carrageenan to yield κ-neocarratetraose sulphate and κ-neocarrahexaose sulphate (Potin et al., 1991). The gene encoding for this κ-carrageenase has been cloned (Barbeyron et al., 1998). The protein is composed of 545 amino acids (Fig. 3) and after removal of a 35 residues signal peptide, has a molecular mass of 57.4 kDa, higher than the molecular mass determined by SDS-PAGE. Electron spray mass spectrometry analysis of the protein purified resulted in a molecular mass of 32,618±5 Da (Barbeyron et al., 1998). Here also, these observations suggest a post-translational removal of a large portion of the C-terminal end of the protein, probably to allow the secretion in outer medium. The cleavage site of this second processing has been determined between the Thr316 and Ser317 (Barbeyron et al., 1998).

Comparison of the two κ-carrageenases with other glycoside hydrolases

The two κ-carrageenases display 36% of identity and 52% of similarity (Fig. 3). Interestingly, *Pseudoalteromonas carrageenovora* and *Zobellia galactanovorans* κ-carrageenase sequences strongly diverge after the C-terminal cleavage site positions. We observe a well conserved segment (EIDVVE) similar to a segment present in the β-agarase proteins from *Z. galactanovorans*. The comparisons between a κ-carrageenase and a β-agarase is always below 25%. Moreover, alignments of a κ-carrageenase with an other members of the family sixteen of glycoside hydrolases show identities often below 20%. To verify if the two κ-carrageenases belong to the family sixteen of glycoside hydrolases, sequence comparisons with a laminarinase, a lichenase and a xyloglucan endotransferase (enzymes which belong to this structural family) have been made, at the secondary structure level, using the hydrophobic cluster analysis method (Fig. 2) (Lemesle-Varloot et al., 1990; Callebaut et al., 1997).

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. Thirteen distinct structural segments (I-XIII) are apparent in the HCA plots (Fig. 2), among which segments I, II, IX, X, XI and XIII appear to be best conserved. 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 2,

Table 2. Similarity scores [a] for pairwise sequence comparisons of family 16 glycoside hydrolases

Tableau 2. Score de similitude [a] des séquences, prises deux à deux, de la famille 16 des glycoside hydrolases.

Sequences [b]	KapPc	KapZg	AgarSc	LamRm	LichBs	XetAt
KapPc	100	71	71	71	72	66
	100	*45*	*24*	*19*	*25*	*20*
KapZg		100	62	67	67	61
		100	*17*	*21*	*18*	*15*
AgarSc			100	61	65	64
			100	*20*	*14*	*18*
LamRm				100	65	71
				100	*12*	*13*
LichBs					100	71
					100	*26*
XetAt						100
						100

[a] For each entry, the average HCA (Hydrophobic Cluster Analysis) score is given on top 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 = 2CR/(RC^1 + RC^2) X 100%, where RC^1 and RC^2 are the numbers of hydrophobic residues in sequence 1 and 2, respectively, and CR is the number of hydrophobic residues in sequence 1 that are in correspondence with sequence 2. The final HCA score was the average value from all of the hydrophobic clusters along the 13 structural segments delineated in figure 2.

[a] Pour chaque entrée, la moyenne des scores HCA est indiquée sur la ligne supérieure et les scores en structure primaire sont donnés en italique. Les scores HCA sont calculés pour chaque amas de la façon suivante : score HCA= 2CR/(RC^1 + RC^2) X 100 %, où RC^1 and RC^2 sont les nombres de résidus hydrophobes dans la séquence 1 et 2 respectivement et CR le nombre de résidus hydrophobes de la séquence 1 qui est en exacte correspondance dans la séquence 2. Le score HCA final est la somme des scores obtenus pour chaque amas divisé par le nombre d'amas pris en considération (13 amas maximum délimités dans la figure 2).

[b] KapPc : κ-carrageenase of *Pseudoalteromonas carrageenovora*; KapZg : κ-carrageenase of *Zobellia galactanovorans* ; AgarSc : β-agarase of *Streptomyces coelicolor*; LamRm: laminarinase of *Rhodothermus marinus* ; LichBs : lichenase of *Bacillus subtilis*; XetAt: xyloglucan endotransglycosylase of *Arabidopsis thaliana*.

Figure 4. Sequence comparison of ι-carrageenases from *Zobellia galactanovorans* (lower sequence) and *Alteromonas fortis* (upper sequence). Arrows and stars indicate the aspartic acid and glutamic acid residues respectively. The aa sequence underlined was confirmed by microsequencing of the purified ι-carrageenase from *Z. galactanovorans*. Arrow in sequences indicate the signal peptidase cleavage sites. Both sequences show 43.2% of identity and 59% of similarities.

Figure 4. Comparaison de séquence des ι-carraghénases de *Zobellia galactanovorans* (séquence inférieure) et d'*Alteromonas fortis* (séquence supérieure). Les flèches et les étoiles indiquent respectivement les acides aspartiques et glutamiques. La séquence d'acides aminés soulignée était confirmée par microséquençage de la ι-carraghénase de *Z. galactanovorans*. La flèche insérée dans les séquences indiquent les sites de coupure de la signal peptidase. Les deux séquences ont 43,2 % d'identité et 59% de similitude.

were above 60 % (Table 2), a significant similarity threshold in this comparison procedure (Lemesle-Varloot et al., 1990). These observations indicate that the κ-carrageenases from *P. carrageenovora* and *Z. galactanovorans* belong to the family sixteen of glycoside hydrolases.

The ι-carrageenases form a new family of glycoside hydrolases

The ι-carrageenase from Alteromonas fortis

The first ι-carrageenase studied was an enzyme purified from an unidentified marine bacterium referred to as strain # 1 in the Yaphe's bacterial collection (Greer & Yaphe, 1984) called here *Alteromonas fortis* (Potin, 1992). This enzyme was used to elucidate the ι-carrageenan structure (Bellion et al., 1981). It is a protein present in the cell-free supernatant with a molecular mass estimated at 57 kDa (Greer & Yaphe, 1984). The ι-carrageenan is cleaved to ι-neocarratetraose sulphate as major end product. The gene of the ι-carrageenase from *A. fortis* has been cloned and its protein sequence determined (Barbeyron et al., 2000). This enzyme is composed of 491 amino acids with a theoretical molecular mass of 54.80 kDa. Here also, the N-terminus of the protein stands out as a domain which represents a signal peptide. The putative cleavage site of the signal peptidase I would be between Gly26 and Ala27 residues. After signal peptide cleavage, the molecular weight of the mature protein shifts down to 51.95 kDa, consistent with the result found from the purified protein (Greer & Yaphe, 1984). It does not seem that there is a second processing allowing the secretion. However, we do not refute the possibility that such a processing exists. It would remove some residues at the N-terminal end. Indeed, all degradation by exonuclease III of the 3'-terminal end of the DNA encoding to the C-terminal of the ι-carrageenase from *Z. galactanovorans* resulted in a loss of activity. Moreover, we observe that there are 19 amino acids between the cutting site of the signal peptidase and the microsequenced N-terminal end of the mature protein of *Z. galactanovorans* (underlined in Fig. 4). This putative processing would allow the release of the enzyme in the extracellular medium.

The ι-carrageenase from Zobellia galactanovorans

A ι-carrageenase activity has been purified from *Z.*

galactanovorans strain Dsij and its encoding gene have been cloned (Barbeyron et al., 2000). The ι-carrageenase from *Z. galactanovorans* displayed an apparent molecular mass of 50 kDa as estimated by SDS-PAGE. Although the ι-carrageenase has been purified from the *Z. galactanovorans* culture supernatant, analysis of its signal peptide suggests that this enzyme might be a lipoprotein. As the agarase B from the same bacterium, the mature protein begins with a cysteine residue (Fig. 4) with a lipidic moiety that could function as a transmembrane anchor. The absence of a negatively charged amino-acid downstream the cleavage site suggests a localization in the outer membrane (Yamaguchi et al., 1988; Gennitzy & Inouye, 1991).

The major end products of degradation of ι-carrageenan, by purified ι-carrageenase from *Z. galactanovorans,* have been identified as ι-neocarratetraose sulphate and the ι-neocarrahexaose sulphate, respectively (Barbeyron et al., 2000). This result is similar to the mode of action of the ι-carrageenase from *Alteromonas fortis* described by Greer and Yaphe (1984). Moreover, ^{1}H NMR analysis of the anomeric configuration of the hydrolysis products of the ι-carrageenase from *A. fortis* indicates that this enzyme inverts the anomeric bond configuration, producing α-anomers that progressively give rise to β-anomers when mutarotation takes place (Barbeyron et al., 2000). It follows that *Z. galactanovorans* and *A. fortis* ι-carrageenases both hydrolyse the β-1,4 linkages of ι-carrageenan by a one-step nucleophilic substitution, a mechanism which results in the inversion of the anomeric configuration and precludes transglycosylation. This finding contrasts with the case of κ-carrageenases, which are retaining glycoside hydrolases with transglycosylating properties (Potin et al., 1995).

The ι-carrageenases are structurally unrelated to κ-carrageenases

After removal of the signal peptide in both sequences, the ι-carrageenase from *Zobellia galactanovorans* was found to share some sequence similarities with ι-carrageenase from *Alteromonas fortis* (Fig. 4). The ι-carrageenase comparison indicates 43.2% of identity and 59% of similarity. The three best conserved domains are in the central region of both sequences from Lys182 to Val200, from Gly227 to Leu259 and Val286 to Met293 (numbering of the *Z. galactanovorans* sequence). We have indicated the invariant Asp and Glu residues, which are good candidates for acting as catalytic residues in glycoside hydrolases (Zvelebil et al., 1988; Henrissat et al., 1989). Only nine such aspartic acid residues were found including three localized in the best conserved regions. In the same way, only five glutamic acid residues are strictly conserved between both sequences, including three in the best conserved regions. These sequence similitudes are emphasized by comparison with HCA method (data not shown) indicating that both ι-carrageenases are related. ι-Carrageenan differs from κ-carrageenan only by the presence of an additional sulphate group on the digalactose repeating unit, at the C_2 position of the α(1,4)-linked galactose residues (Fig. 1). Yet, we show here that the ι-carrageenases are totally unrelated to κ-carrageenases in their amino acid sequences, and in consequence most likely also at the three-dimensional level. Especially the ι-carrageenases lack the motif E[ILV]D[IVAF][VILMF](0,1)E, i.e. the catalytic site typical of clan B of glycoside hydrolases (Planas et al., 1992; Juncosa et al., 1994; Krah et al., 1998), which encompasses glycoside hydrolase families 7 and 16 (Davies & Henrissat, 1995).

The α-agarase from Alteromonas agarilytica *is an unclassified glycoside hydrolase*

The properties of the Alteromonas agarilytica *α-agarase*

A number of micro-organisms were reported to use agarose as a carbon source. For example, we already mentioned agarases from *Pseudolateromonas atlantica*, *Streptomyces coelicolor* (Buttner et al., 1987), and *Zobellia galactanovorans* (Flament et al., unpublished). All enzymes cleave the β-1,4 linkage in agarose. A marine bacterium, referred to as strain GJ1B, produced an agarase which cleave the α-1,3 linkage in the agarose (Young et al., 1978), yielding oligosaccharides of agarobiose series, i.e., with 3,6-anhydro-L-galactose residues at the reducing ends and D-galactose residues at the non-reducing ends. This strain has been described (Young et al., 1978; Potin et al., 1993) and tentatively assigned to the species *Alteromonas agarilytica* Strain GJ1B (Potin et al., 1993). The purified protein is a dimer, with a molecular mass of 360 kDa, which may be dissociated in two active subunits of 180 kDa as determined by SDS-PAGE (Potin et al., 1993). The major end products of enzymatic hydrolysis of the agarose by the α-agarase are the agarohexaose and the agarotetraose (Young et al., 1978). The latter is further cleaved in agarotriose and galactose by an associated β-galactosidase activity with the α-agarase (Potin et al., 1993).

Organization of the α-agarase protein

The *Alteromonas agarilytica* α-agarase gene has been cloned (Flament et al., unpublished). The protein sequence is composed of 1429 amino acids (GenPept: AAF26838.1). Analysis of the N-terminus of the protein suggests that a signal peptide is present with a most probable cleavage site for the signal peptidase I between Ala26 and Glu27. The theoretical molecular mass of the protein after removal of the signal peptide, is 151.93 kDa. The protein sequence compared with proteins present in the data bank shows that

the α-agarase is a bimodule protein with a N-terminal domain and a C-terminal domain of 800 and 600 residues respectively.

The N-terminal domain appears composed of two types of repeating elements. Immediately downstream the signal peptide, the amino acid sequence (noted A in Fig. 5) is similar to the carbohydrate-binding modules (CBMs) family 6, present in cellulase B (CelB) from *Cellvibrio mixtus* (Fontes et al., 1998). Two others (noted B and C in Fig. 5) ca. 130 aa-long CBMs are found. In addition, five sequences (numbered I to V in Fig. 5) with a well conserved pattern of 33 amino acids (identity scores above 55%) are reiterated. These repeats resemble the calcium-binding sites, referred to as type-3 repeats of thrombospondin, an adhesive protein that mediates intercellular interactions in animal cells (Hynes, 1985; Lawler & Hynes, 1986). The presence, in α-agarase from *Alteromonas agarilytica*, of repeated sequences similar to the type-3 calcium-binding sites of thrombospondin is consistent with the observations of Young et al. (1978) and Potin et al. (1993) showing that α-agarase activity is stabilized by the presence of calcium ions.

The C-terminal domain does not present any similarity with other known proteins. It is possible that the catalytic site is included in this domain. The α-agarase from *A. agarilytica* is, to date, an unclassified glycoside hydrolase.

Structural biology of galactan hydrolases

The family sixteen of glycoside hydrolases

The main structural characteristics of agarases and carrageenases are summarized in Table 3 and figure 5. Upon phylogenetical analysis of family 16, glycoside hydrolases cluster according to their substrate specificity (Fig. 6). Like the lichenases and xyloendoglucantransferases (XET), β-agarases and κ-carrageenases are structurally homologous to family-16 laminarinases, from which they have probably evolved by gene duplication, as arose the need for degrading novel polysaccharidic motives in the evolution of eubacteria (Barbeyron et al., 1998). They have retained the same hydrolytic mechanism, general folding and catalytic sites, which can accommodate for the presence of zero or one sulphate substituent on the digalactose repeating units of agars and carrageenans (Fig. 1). This finding may be accounted for by the difference between agars and

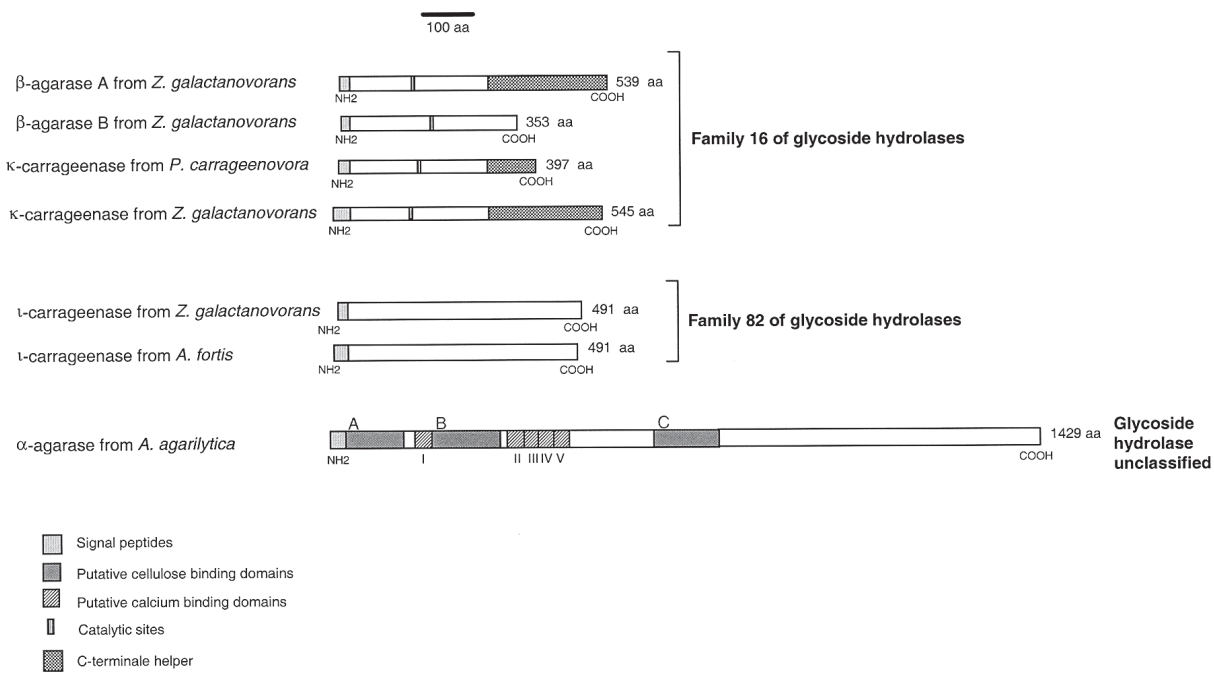
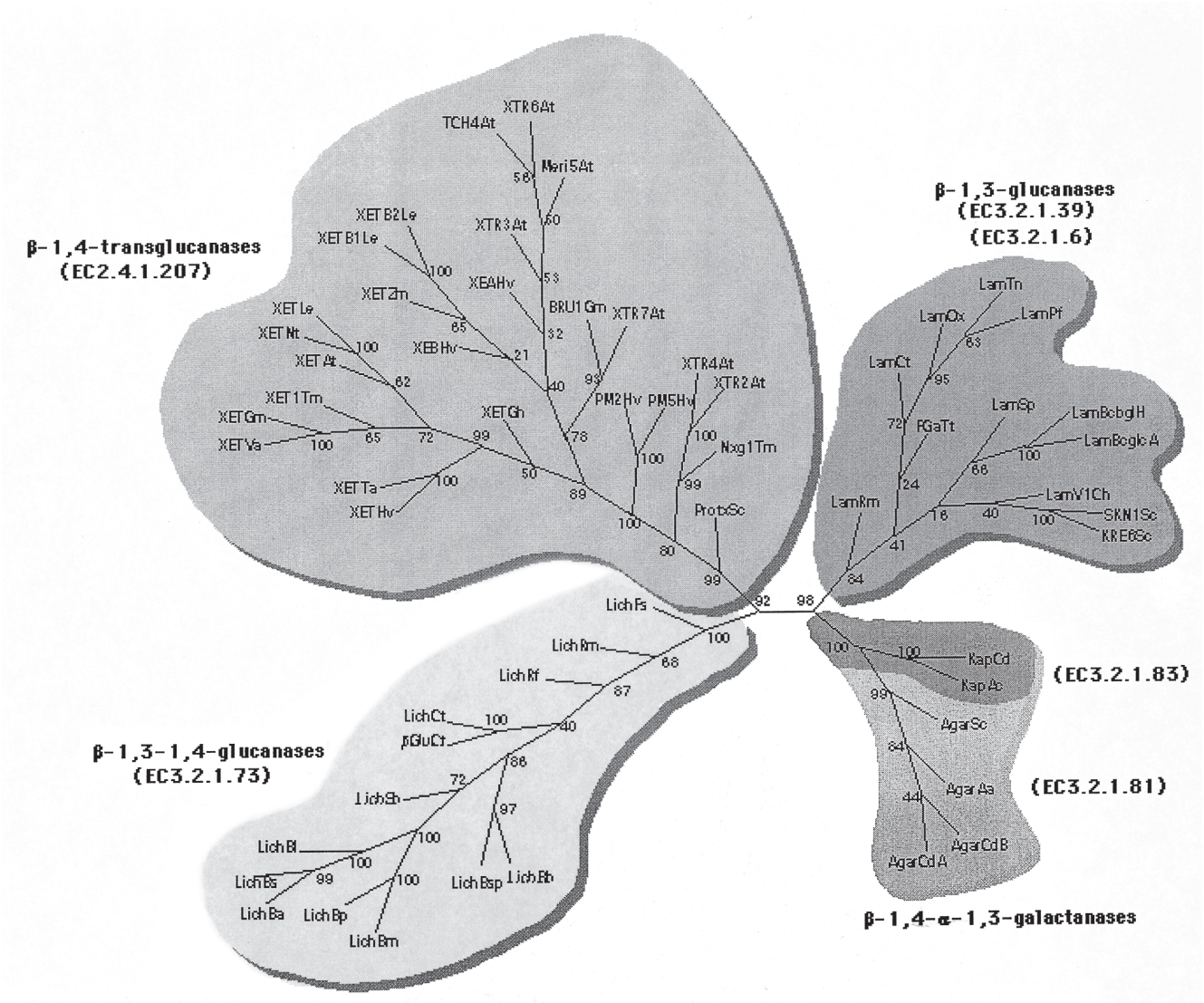


Figure 5. Schematic representation of galactan hydrolases. For each enzyme different domains known are displayed and their respective structural family classification is given.

Figure 5. Représentation schématique des galactane hydrolases. Pour chaque enzyme les différents domaines connus sont montrés ainsi que leur appartenance aux différentes familles structurales de glycoside hydrolases.



κ-carrageenans in the D/L isomery of the α(1,4)-linked galactose residues. As a member of family-16 glycoside hydrolases, β-agarases and κ-carrageenases should display a general folding similar to the 3-D structures known for *Bacillus licheniformis* and *B. macerans* lichenases (Keitel et al., 1993; Hahn et al., 1995). However, in spite of their structural relationships in hydrophobic cluster analysis, β-agarases or κ-carrageenases share with lichenases only, a 20% overall identity in their primary structures, far below the threshold to allow modelisation experiments. Moreover, in the conserved catalytic β-stranded domain, β-agarases and κ-carrageenases (as well as laminarinases) depart from lichenases (and XETs) by the presence of one additional amino-acid, an insertion which would affect the active site fine topology (Fig. 7). To obtain a high-resolution three-dimensional structure of a κ-carrageenase, we have therefore overexpressed in *Escherichia coli* of large amount of recombinant enzyme from *P. carrageenovora* (Michel et al., 1999). Crystallization experiments have given crystals (Fig. 8a). Crystals of *P. carrageenovora* κ-carrageenase diffracted up to 2.4 Å and were very stable in the X-ray beam. These crystals belong to the orthorhombic space group $P2_12_12_1$, with unit-cell dimensions of: a = 58.2 Å, b = 62.8 Å, c = 77.9 Å, α = β = γ = 90° (Michel et al., 1999). Assuming the presence of one molecule of recombinant κ-carrageenase in the asymmetric unit, resulted in a calculated molecular volume (Vm) of 2.16 $Å^3$ Da^c and a corresponding solvent content of 37%, consistent with the range frequently observed for protein crystals (Matthews, 1968). The κ-carrageenase tridimensional structure was determined at 1.54 Å using the multiwavelength anomalous diffraction (MAD) method with a crystal labelled by seleno-methionine (Michel et al., 2001). The asymmetric unit encompasses aminoacids 27-297 of κ-carrageenase, 7

Figure 6. Parsimony phylogenetic tree of the family-16 glycoside hydrolases, including various xyloglucan endotransglycosylases, laminarinases, lichenases and galactanases. Numbers indicate the bootstrap values in the parsimony analysis. The distance matrix tree (not shown) was similar to the parsimony tree.

Figure 6. Arbre phylogénétique de parsimonie de la famille 16 des glycoside hydrolases, comprenant différentes xyloglucan endotransglycosylases, laminarinases, lichénases et galactanases. Les nombres indiquent les valeurs de bootstrap dans l'analyse de parsimonie. L'arbre de distance obtenu par la méthode du plus proche voisin, en utilisant la matrice de Dayoff (non montré) était semblable à celui obtenu par parsimonie.

LamRm: laminarinase of *Rhodothermus marinus* (SWISS-PROT U04836); **LamBcglca**: *glcA* laminarinase gene of *Bacillus circulans* (SWISS-PROT P23903); **LamBcbglh**: *bglH* laminarinase gene of *B. circulans* (DDBJ D17519); **LamTn**: laminarinase of *Thermotoga neapolitina* (EMBL Z47974); **LamCt**: laminarinase of *Clostridium thermocellum* (EMBL X89732); **LamV1Ch**: laminarinase of *Paramecium bursaria Chlorella* virus1 (NCBI 1131438); **LamSp**: laminarinase of *Strongylocentrotus purpuratus* (GenBank U49711); **LamOx**: laminarinase of *Oerskovia xanthineolytica* (GenBank U56935); **LamPf**: 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* (PIR A41624); **SKN1Sc**: β-glucan synthesis-associated protein of *Saccharomyces cerevisiae* (SWISS-PROT P33336). **LichBm**: lichenase of *Bacillus macerans* (SWISS-PROT P23904); **LichBp**: lichenase of *Bacillus polymyxa* (SWISS-PROT P45797); **LichBa**: lichenase of *Bacillus amyloliquefaciens* (SWISS-PROT P07980); **LichBs**: lichenase of *Bacillus subtilis* (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 296932); **LichCt**: 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); **LichRm**: lichenase of *Rhyzobium meliloti* (SWISS-PROT P33693); **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); **XETNt**: XET of *Nicotiana tobacum* (DDBJ D86730); **XETHv**: XET of *Hordeum vulgare* (EMBL X91659); **XEAHv**: XET like protein of *Hordeum vulgare* (EMBL X93174); **XEBHv**: XET like protein of *Hordeum vulgare* (EMBL X93175); **PM2Hv**: XET like protein of *Hordeum vulgare* (EMBL X91660); **PM5Hv**: XET like protein of *Hordeum vulgare* (EMBL X93173); **XET1Tm**: XET from epicotyls of *Tropaelum majus* (GenBank L43094); **Nxg1Tm**: XET from seeds of *T. majus* (GenBank X68254); **TCH4At**: XET of *A. thaliana* (GenBank U27609); **XTR2At**: xyloglucan endotransglycosylase-related protein of of *A. thaliana* (GenBank U43487); **XTR3At**: xyloglucan endotransglycosylase-related protein of of *A. thaliana* (GenBank U43485); **XTR4At**: xyloglucan endotransglycosylase-related protein of of *A. thaliana* (GenBank U43486); **XTR6At**: xyloglucan endotransglycosylase-related protein of of *A. thaliana* (GenBank 43488); **XTR7At**: xyloglucan endotransglycosylase-related protein of of *A. thaliana* (GenBank U43489); **ProtxSc**: hypothetical protein L8167.22 of *Saccharomyces cerevisiae* (PIR S48564). **AgarSc**: β-agarase of *Streptomyces coelicolor* (SWISS-PROT P07883); **AgarPa**: β-agarase of *Pseudoalteromonas atlantica* (GeneBank M73783); **AgarACd**: β-agarase A of *Cytophaga drobachiensis* (*Zobellia galactanovorans* ; GenPept AAF21820.1) and **AgarBCd**: β-agarase B of *C. drobachiensis* (*Zobellia galactanovorans* ; GenPept AAF21821.1); **κKapPc**: β-carrageenase of *P. carrageenovora* (SWISS-PROT P43478); **KapCd**: κ-carrageenase of *C. drobachiensis* (*Zobellia galactanovorans* ; SWISS-PROT O84907).

cadmium ions, 7 chloride anions and 405 water molecules. κ-Carrageenase is an almost all-β protein with a globular shape of 60 x 50 x 40 Å in dimension, folded in a curved β-sandwich of two antiparallel β-sheets composed of 6 and 7 strands respectively (Fig. 8b). It results in the presence of a tunnel-like active site cavity (Michel et al., 2001). As expected from the secondary structure predictions using HCA method (Barbeyron et al., 1994, 1998), the fold of κ-carrageenase from *P. carrageenovora* is closely related to enzymes of the family-16 hydrolases. The catalytic residues E163 (nucleophile residue) and E168 (acid/base residue) are, as expected for the retaining glycoside hydrolases, separated by 5.4 Å. Compared to the homologous catalytic residues in the lichenase from *Bacillus macerans*, a hydrophobic aminoacid is inserted in the catalytic site of the κ-carrageenase (Fig. 7). However, in spite of this insertion, the catalytic aminoacids of both enzymes are exactly superimposed due to the presence, in the κ-carrageenase, of a β-bulge absent in the lichenase. The crystal structure of the *P. carrageenovora* carragenase is the first described tridimensional structure of a carrageenase. Its tunnel-shape active site, the first to be reported for enzymes other than cellulases, suggests that such tunnels are associated with the degradation of solid polysaccharides (Michel et al., 2001). The small size of the catalytic tunnel κ-carrageenase cannot contain a double helix of κ-carrageenan present in the gel. It follows that the enzyme hydrolyses κ-carrageenan chains either as single-stranded helices or as random coils. The occurrence of this catalytic tunnel, rather than a deep channel as observed in lichenase of the family 16 of glycoside hydrolases from *Bacillus macerans*, implies a processive mechanism of hydrolysis of κ-carrageenan. This finding raises the question of how, in such an endo-processive enzyme, occurs the initial attack. As recently demonstrated for *Humicola insolens* (Armand et al., 1997) and *Trichoderma reesei* (Zou & Jones, 1999) cellobiohydrolases, the initial endo-attack by κ-carrageenase could involve conformational changes opening the tunnel-forming loops.

Table 3. Molecular features of galactan hydrolases
Tableau 3. Caractéristiques moléculaires des galactane hydrolases

β-Agarases	**AgaA Zg**	**AgaB Zg**
End product	Neoagarotetraose	Neoagarotetraose
Preproprotein	59.98 kDa	40.68 kDa
Mature protein	31 kDa	?
C-terminal cleavage	Yes	No
Hydrolytic mechanism	Retention	Retention
Transglycosylating activity	Yes	Yes
Structural family	16	16

κ-carrageenases	**CgkA Pc**	**CgkA Zg**
End product	κ-Neocarrabiose and tetraose	κ-Neocarratetraose
Preproprotein	44.2 kDa	61.4 kDa
Mature protein	31.7 kDa	32.6 kDa
C-terminal cleavage	N ^{301}	T ^{316}
Hydrolytic mechanism	Retention	Retention
Transglycosylating activity	Yes	Yes
Structural family	16	16

ι-carrageenases	**CgiA Af**	**CgiA Zg**
End product	ι-Neocarratetraose	ι-Neocarratetraose
Preprotein	54.8 kDa	53.4 kDa
Mature protein	55 kDa	50 kDa
C-terminal cleavage	No	No
Hydrolytic mechanism	Inversion	Inversion
Transglycosylating activity	No	No
Structural family	New	New

α-agarase	**AgaA Aa**
End product	Agarotetraose and agarohexaose
Preprotein	154.69 kDa
Mature protein	151.93 kDa
C-terminal cleavage	No
Hydrolytic mechanism	unknown
Transglycosylating activity	unknown
Structural family	unclassified

Pc: *Pseudoalteromonas carrageenovora*; Af: *Alteromonas fortis*; Zg: *Zobellia galactanovorans*; Aa: *Alteromonas agarilytica*.

The ι-carrageenases from Alteromonas fortis *and* Zobellia galactanovorans

As shown in figure 1, ι-carrageenans differ from κ-carrageenans only by the presence of an additional sulphate group on the digalactose repeating unit, at the C_2 position of the α(1,4)-linked galactose residues. Yet ι-carrageenases are totally unrelated to κ-carrageenases in their amino acid sequence, including the three-dimensional level and their hydrolytic mechanism. The fact that different proteins are required to hydrolyse κ- and ι-carrageenans suggests that the number (and position) of sulphate ester substituents on the polygalactose backbone represents a constraint in the evolution of galactan hydrolases. In this context, we are looking forward to the characterization of the structural gene of λ-carrageenase which hydrolyse a carrageenan with three sulphate substituents by repeating unit (Fig. 1) and that may elucidate the structure of a novel galactan hydrolase.

The different mechanisms observed between ι- and κ-carrageenases together with their lack of sequence similarities imply a possible different fold. A high-resolution three-dimensional structure of a ι-carrageenase was therefore required for defining amino-acid residues involved in the recognition and cleavage of ι-carrageenans. After overproduction of the *A. fortis* ι-carrageenase in *E. coli,* we have obtained ι-carrageenase crystals (Michel et al., 2000). The crystals belong to the monoclinic space group $P2^1$, with unit-cell parameters : a = 90.1 Å, b = 122.0 Å, c = 58.2 Å, α = γ = 90°, β = 91.3°. Assuming the presence of two molecules of recombinant ι-carrageenase in the asymmetric unit resulted in a calculated molecular volume (Vm) of 3.0 $Å^3$ Da^{-1} consistent with the range frequently observed for protein crystals (Matthews, 1968).

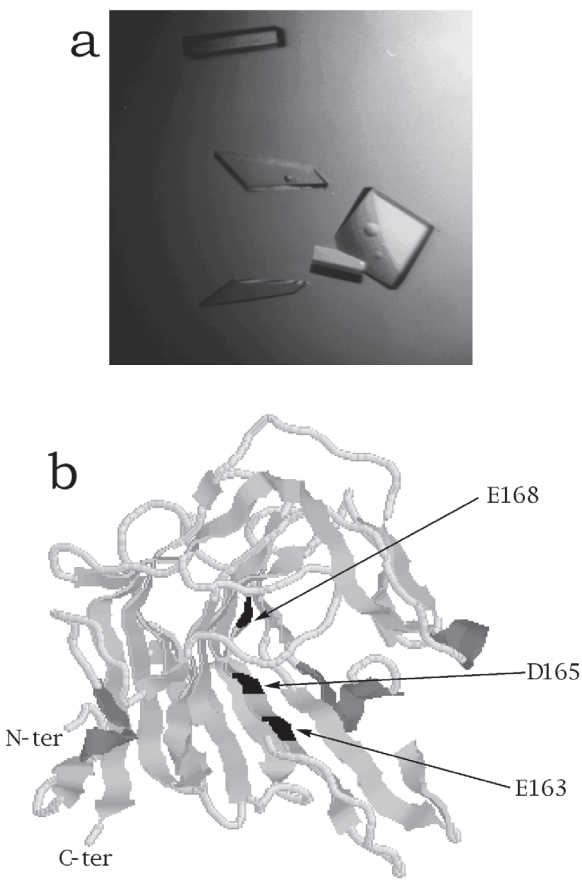
The Alteromonas agarilytica *α-agarase*

The α-agarase from *A. agarilytica* shares no structural relationship with β-agarases, i.e., the hydrolases that degrade agar by splitting the β-1,4, equatorial linkages of the polygalactose backbone. In particular, it is not structurally related to the β-agarase from *Streptomyces coelicolor* as well as to those of *Pseudoalteromonas atlantica* and *Zobellia galactanovorans*, which all belong to family 16 of glycoside hydrolases (Flament et al., unpublished results). *A. agarilytica* α-agarase does not show any homology with the β-agarases A and B from *Vibrio* sp. strain JT0107 (Sugano et al., 1993; 1994), which belong to family 50 of glycoside hydrolases (Henrissat & Bairoch, 1996). Therefore, agarases have evolved from at least three different progenitor sequences, leading to extant family-16 and family-50 β-agarases as well as α-agarase, respectively. In the case of α-agarase or at least its N-terminal moiety, the ancestral protein could be a cellulase or a xylanase. Analysis of the α-agarase sequence suggests that β- and α-agarases are markedly different in both their overall organization and the secondary structure of their catalytic sites. We know that

β-1,3-glucanases			
	LamRm	EIDIME	*Rhodothermus marinus* (*Bacteria*, *Proteobacteriae*)
	FGaTt	EIDFIE	*Tachypleus tridentatus* (Arthropod)
	LamSp	EIDLVE	*Strongylocentrotus purpuratus* (Echinoderm, sea urchin)
	Kre6Sc	EIDVLE	*Saccharomyces cerevisiae* (Fungi)
	LamPf	EIDIME	*Pyrococcus furiosus* (*Archae,* hyperthermophile)
	LamV1Ch	EIDIFE	Virus de Chlorella
	LamTn	EIDIME	*Thermotoga neapolitana* (*Bacteria*, *Thermotogales*)
β-1,4-α-1,3-galactanases			
	AgarPa	EIDAME	*Pseudoalteromonas atlantica* (*Bacteria*, *Proteobacteria* γ)
	AgarSc	EIDVIE	*Streptomyces coelicolor* (*Bacteria*, *Actinomycetales*)
	AgarBZg	EIDILE	*Zobellia galactanovorans* (*Bacteria*, *Cytophagales*)
	AgarAZg	EIDIME	*Zobellia galactanovorans* (*Bacteria*, *Cytophagales*)
	KapZg	EIDVVE	*Zobellia galactanovorans* (*Bacteria*, *Cytophagales*)
	KapPc	EIDVVE	*Pseudoalteromonas carrageenovora* (*Bacteria*, *Proteobacteria* γ)
β-1,3-1,4-glucanases			
	LichBm	EIDI-E	*Bacillus macerans* (*Bacteria* Gram positive)
	LichBl	EIDI-E	*Bacillus licheniformis* (*Bacteria* Gram positive)
β-1,4-transglucanases			
	XetAt	EIDF-E	*Arabidopsis thaliana* (Plant dicot)
	XetHv	EIDF-E	*Hordeum vulgare* (Plant monocot)

Figure 7. Primary structure comparison of the catalytic sites from various members of family-16 glycoside hydrolases. The catalytic glutamic acids are boxed. Note that there is an amino acid in less in lichenase and xyloglucan endotransglycosylase sequences.

Figure 7. Comparaison des structures primaires des sites catalytiques de différents membres de la famille 16 des glycosides hydrolases. Les acides glutamiques catalytiques sont entourés. Notez l'absence d'un acide aminé dans le site catalytique des lichénases et des xyloglucane endotransglycosidases.



the κ-carrageenases and β-agarases belong in the same structural family of glycoside hydrolases even though their substrates are different (divergent evolution). By contrast, but with the limitation that, to date, the primary structure of only one α-galactanase is known, this work shows that completely unrelated glycoside hydrolases are required to cleave either one or the other linkage type of a same

Figure 8. A. κ-Carrageenase crystals were obtained at 15°C using the hanging-drop vapour diffusion method with polyethylene glycol (MW 4,000) as precipitating agent. **B.** Folding of the *P. carrageenovora* κ-carrageenase, view of the ribbon representation of the structure resolved from the above crystals. κ-Carrageenase is an almost all-β protein with a globular shape. The catalytic amino acids (E163, E168) and an aspartic acid (D165) essential for catalysis are pointed by arrows inside the catalytic groove.

Figure 8. A. Les cristaux de κ-carraghénase ont été obtenus à 15°C par la méthode de diffusion de vapeur et le polyéthylène glycol (MW 4000) comme agent précipitant. **B.** Structure tridimensionnelle de la κ-carraghénase de *P. carrageenovora*. Représentation en rubans de la structure résolue à partir des cristaux présentés ci-dessus. La κ-carraghénase est une protéine essentiellement composée de feuillets β de forme globulaire. Les acides aminés catalytiques (E163, E168) and un acide aspartique (D165) essentiel pour la catalyse sont indiqués par des flèches dans la cavité du site catalytique.

substrate (convergent evolution) with alternating axial and equatorial glycosydic bonds (agarose). This conclusion suggests that the linkage type (α or β) in agarose has constrained the evolution of agarases.

Acknowledgements

This research was supported by the priority programmes ACC "Biologie Structurale" and GDR 1002 (CNRS-Ifremer), as well as by research grants from the Laboratoires Goëmar, St Malo, France.

References

Armand S., Drouillard S., Schulein M., Henrissat B. & Driguez H. 1997. A bifunctionalized fluorogenic tetrasaccharide as a substrate to study cellulases. *Journal of Biological Chemistry,* **272**: 2709-2713.

Barbeyron T. 1993. *Caractérisation moléculaire d'une sulfatase, d'une kappa-carraghénase et d'une iota-carraghénase chez deux bactéries marines :* Alteromonas carrageenovora *et* Cytophaga drobachiensis. Thèse Doctorat Université Pierre & Marie Curie-Paris VI. 260 pp.

Barbeyron T., L'Haridon S., Corre E., Kloareg B., & Potin P. 2001. *Zobellia galactanovora* gen. nov., sp. nov, a marine species of Flavobacteriacae isolated from a red alga, and classification of *[Cytophaga] uliginosa* (ZoBell & Upham, 1944) Reichenbach 1989 as *Zobellia uliginosa* gen. nov., comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, **51**: in press.

Barbeyron T., Michel G., Potin P., Henrissat B. & Kloareg B. 2000. ι-Carrageenase of the marine bacterium *Cytophaga drobachiensis*. *Journal of Biological Chemistry*, **15**: 528-537.

Barbeyron T., Gérard A., Potin P., Henrissat B. & Kloareg B. 1998. The kappa-carrageenase of the marine bacterium *Cytophaga drobachiensis*: Structural and phylogenetic relationships within family-16 glycoside hydrolases. *Molecular Biology and Evolution*, **15**: 528-537.

Barbeyron T., Henrissat B. & Kloareg B. 1994. The gene encoding the kappa-carrageenase of *Alteromonas carrageenovora* is related to β-1,3-1,4-glucanases. *Gene,* **139**: 105-109.

Bellion C., Hamer G.K. & Yaphe W. 1981. Analysis of kappa-iota hybrid carrageenans with kappa-carrageenase, iota-carrageenase and 13C-NMR. *Proceedings of the International Seaweed Symposium*, **10**: 379-384.

Buttner M.J., Fernley I.M. & Bibb M.J. 1987. The agarase gene *dag A* of *Streptomyces coelicolor* A32: nucleotide sequence and transcriptionnal analysis. *Molecular and General Genetics*, **209**: 101-109.

Callebaut I., Labesse G., Durand P., Poupon A., Canard L., Chomilier J., Henrissat B. & Mornon J.-P. 1997. Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cellular and Molecular Life Sciences*, **53**: 621-645.

Carlucci M.J., Pujol C.A., Ciancia M., Noseda M.D., Matulewiicz M.C., Diamonte E.B. & Cerezo A.S. 1997. Antiherpetic and anticoagulant properties of carrageenans from the red seaweed *Gigartina skottsbergii* and their cyclized derivatives: correlation between structure and biological activity. *International Journal of Biological Macromolecules*, **20**: 97-105.

Davies G. & Henrissat B. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure*, **3:** 853-859.

De Ruiter G.A. & Rudolph B. 1997. Carrageenan biotechnology: an industrial perspective. In: *Marine microorganisms: Research issues for biotechnology*, pp. 187-188. Edition European Community and United States workshop: Bruxelle, Belgium.

Di Rosa M. 1972. Biological properties of carrageenan. *Journal of Pharmacy and Pharmacology*, **24**: 89-102.

Fontes C.M., Clarke J.H., Hazlewood G.P., Fernandes T.H., Gilbert H.J. & Ferreira L.M. 1998. Identification of tandemly repeated type VI cellulose-binding domains in an endoglucanase from the aerobic soil bacterium *Cellvibrio mixtus*. *Applied Microbiology and Biotechnology*, **495**: 552-559.

Greer C.W. & Yaphe W. 1984. Purification and properties of ι-carrageenase from a marine bacteria. *Canadian Journal of Microbiology*, **30**: 1500-1506.

Henrissat B. & Bairoch A. 1996. Updating the sequence-based classification of glycosyl hydrolases. *Biochemical Journal*, **316**: 695-696.

Henrissat B., Claeyssens M., Tomme P., Lemesle L. & Mornon J.-P. 1989. Cellulase families revealed by Hydrophobic Cluster Analysis. *Gene*, **81**: 83-95.

Hoffman R. 1993. Carrageenans inhibit growth-factor binding. *Biochemical Journal,* **289**: 331-334.

Høj P. B., Condron R., Traeger J. C., McAuliffe J. C. & Stone B. A. 1992. Identification of glutamic acid 105 at the active site of *Bacillus amyloliquefaciens* 1,3-1,4-β-D-glucan 4-glucanohydrolase using epoxide-based inhibitors. *Journal of Biological Chemistry,* **267**: 25059-25066.

Hynes R. 1985. Molecular biology of fibronectin. *Annual Review of Cellular Biology*, **1**: 67-90.

Juncosa M., Pons J., Dot T., Querol E. & Planas A. 1994. Identification of active site carboxylic residues in *Bacillus licheniformis* 1,3-1,4-β-D-glucan-4-glucanohydrolase by site directed mutagenesis. *Journal of Biological Chemistry,* **269**: 14530-14535.

Kloareg B. & Quatrano R.S. 1988. Structure of the cell walls of marine algae and ecophysiological functions of the matrix polysaccharides. *Oceanography and Marine Biology an Annual Review*, **26**: 259-315.

Krah M., Misselwitz R., Politz O., Thomsen K. K., Welfle H. & Borriss R. 1998. The laminarinase from thermophilic eubacterium *Rhodothermus marinus*. Conformation, stability, and identification of active site carboxylic residues by site directed mutagenesis. *European Journal of Biochemistry,* **257**: 101-111.

Lawler J. & Hynes R.O. 1986. The structure of human thrombospondin, an adhesive glycoprotein with multiple calcium-binding sites and homologies with several different proteins. *Journal of Cellular Biology*, **1035**: 1635-1648.

Lemesle-Varloot L., Henrissat B., Gaboriaud C., Bissery V., Morgat A. & Mornon J.-P. 1990. Hydophobic cluster analysis: procedures to derive structural and functional information from 2D-representation of protein sequences. *Biochimie*, **72**: 555-574.

Matthews B.W. 1968. Solvent content of protein crystals. *Journal of Molecular Biology*, **33**: 491-497.

McLean M.W. & Williamson F.B. 1979. Kappa-carrageenase from *Pseudomonas carrageenovora*. *European Journal of Biochemistry*, **93**: 553-558.

McLean M.W. & Williamson F.B. 1981. Enzymes from *Pseudomonas carrageenovora*. Application to studies of carrageenan structure. *Proceeding International Seaweed Symposium*, **10**: 479-484.

Michel G., Barbeyron T., Flament D., Vernet T., Kloareg B. & Dideberg O. 1999. Expression, purification, crystallisation and preliminary X-ray analysis of the κ-carrageenase from *Pseudoalteromonas carrageenovora*. *Acta Crystallographica*, *Section D*, **55** : 918-920.

Michel G., Barbeyron T., Flament D., Vernet T., Kloareg B. & Dideberg O. 2000. Expression, purification, crystallisation and preliminary X-ray analysis of the ι-carrageenase from *Alteromonas fortis*. *Acta Crystallographica*, *Section D*, **56** : 766-768.

Michel G., Chantalat L., Duée E., Barbeyron T., Henrissat B., Kloareg B. & Dideberg O. 2001. The κ-carrageenase from *Pseudoalteromonas carrageenovora* features a tunnel-shaped active site: a novel insight in the evolution of clan -B glycoside hydrolases. *Structure,* in press.

Morrice L.M., McLean N.W., Long W.F. & Williamson F. B. 1983a. β-agarases I and II from *Pseudomonas atlantica*. Purification and some properties. *European Journal of Biochemistry*, **135**: 553-558.

Morrice L.M., McLean N.W., Long W.F. & Williamson F. B. 1983b. β-agarases I and II from *Pseudomonas atlantica*. Substrate specificities. *European Journal of Biochemistry*, **137**: 149-154.

Patier P., Potin P., Rochas C., Kloareg B., Yvin J.-C. & Liénart Y. 1995. Free and silica-bound oligo kappa-carrageenan elicit laminarinase activity in *Rubus* cells and protoplasts. *Plant Science,* **110**: 27-35.

Planas A., Juncosa M., Lloberas J. & Querol E. 1992. Essential catalytic role of Glu134 in endo-β−1,3-1,4-D-glucan 4-glucanohydrolase from *B. licheniformis* as determined by site directed mutagenesis. *FEBS letter,* **308**: 141-145.

Potin P. 1992. *Recherche, production, purification et caractérisation de galactane-hydrolases pour la préparation d'oligosaccharides des parois d'algues rouges.* Thèse Doctorat Université de Bretagne Occidentale. 186 pp.

Potin P., Richard C., Barbeyron T., Henrissat B., Gey C., Pétillot Y., Forest E., Dideberg O., Rochas C. & Kloareg B. 1995. Processing and hydrolytic mechanism of the *cgkA*-encoded κ-carrageenase of *Alteromonas carrageenovora*. *European Journal of Biochemistry,* **228**: 971-975.

Potin P., Richard C., Rochas C. & Kloareg B. 1993. Purification and characterization of the α-agarase from *Alteromonas agarlyticus* Cataldi comb. nov., strain GJ1B. *European Journal of Biochemistry,* **214**: 599-607.

Potin P., Sanséau A., Le Gall Y., Rochas C. & Kloareg B. 1991. Purification and characterization of a new κ-carrageenase from a marine *Cytophaga*-like bacterium. *European Journal of Biochemistry,* **201**: 241-247.

Rees D.A., 1969. Structure, conformation and mechanism in the formation of polysaccharide gels and networks. *Advances in Carbohydrate Chemistry and Biochemistry,* **24**: 267-332.

Rochas C., Potin P. & Kloareg B. 1994. NMR spectroscopic investigation of agarose oligomers produced by an alpha-agarase. *Carbohydrate Research,* **253**: 69-77.

Rochas C., Rinaudo M. & Vincendon M. 1983. Spectroscopic characterization and conformation of oligo kappa carrageenan. *International Journal of Biological Macromolecules*, **5**: 111-115.

Sugano Y., Matsumoto T. & Noma M. 1994. Sequence analysis of the *agaB* gene encoding a new beta-agarase from *Vibrio* sp. strain JT0107. *Biochimica et Biophysica Acta*, **12181**: 105-108.

Sugano Y., Matsumoto T., Kodama H. & Noma M. 1993. Cloning and sequencing of *agaA*, a unique agarase 0107 gene from a marine bacterium, *Vibrio* sp. strain JT0107. *Applied and Environmental Microbiology*, **5911**: 3750-3756.

Tomme P., Warren R.A. & Gilkes N.R. 1995. Cellulose hydrolysis by bacteria and fungi. *Advances in Microbial Physiology,* **37**: 1-81.

Weigl J. & Yaphe W. 1966. The enzymic hydrolysis of carrageenan by *Pseudomonas carrageenovora*: Purification of a κ-carrageenase. *Canadian Journal of Microbiology*, **12**: 939-947.

Yamada T., Ogamo A., Saito T., Watanabe J., Uchiyama H. & Nakagawa Y. 1997. Preparation and anti-HIV activity of low-molecular-weight carrageenans and their sulfated derivatives. *Carbohydrate Polymers*, **32**: 51-55.

Yamaguchi K., Yu F. & Inouye M. 1988. A single amino acid determinant of the membrane localization of lipoproteins in *E. coli*. *Cell,* **53**: 423-432.

Yaphe W. & Baxter B. 1955. The enzymic hydrolysis of carrageenin. *Applied Microbiology*, **3**: 380-383.

Yoshizawa Y., Ametani A., Tsunehiro J., Nomura K., Itoh M., Fukui F. & Kaminogawa S. 1995. Macrophage stimulation activity of the polysaccharide fraction from a marine alga (*Porphyra yezoensis*): structure-function relationships and improved solubility. *Biosciences, Biotechnology and Biochemistry,* **59**:1933-1937.

Young K.S., Bhattacharjee S.S. & Yaphe W. 1978. Enzymic cleavage of the α-linkages in agarose to yield agarooligosaccharides. *Carbohydrate Research*, **66**: 207-212.

Zou J.Y. & Jones T.A. 1999. Crystallographic evidence for substrate ring distortion and protein conformational changes during catalysis in cellobiohydrolase Cel6A from *Trichoderma reesei*. *Structure*, **7**: 1035-1045.

Zvelebil M.J. & Sternberg M.J. 1988. Analysis and prediction of the location of catalytic residues in enzymes. *Protein Engineering*, **2**: 127-138.