CMLS, Cell. Mol. Life Sci. 57 (2000) 1135–1143
1420-682X/00/0701135-09 $ 1.50 + 0.20/0

CMLS Cellular and Molecular Life Sciences

Research Article

Molecular cloning of a cDNA encoding alpha-glucosidase in the digestive gland of the shrimp, *Litopenaeus vannamei*

P. Le Chevalier[a,*], D. Sellos[b] and A. Van Wormhoudt[b]

[a]Laboratoire Universitaire de Microbiologie Appliquée de Quimper, 6 rue de l'université, 29334 Quimper cedex (France), Fax + 33 298 908544, e-mail: chevalie@iutquimp.univ-brest.fr
[b]Marine Biological Station, Museum National d'Histoire Naturelle et Collège de France, BP 225, 29182 Concarneau cedex (France)

Received 30 March 2000; received after revision 15 May 2000; accepted 31 May 2000

Abstract. The complete sequence of the 3-kb cDNA and the 5′ genomic structure are reported for the gene encoding the shrimp alpha-glucosidase. Alpha-glucosidase cDNA was isolated from a shrimp digestive gland cDNA library. The 2830-base pair cDNA contains an open reading frame that encodes 919 amino acids. The shrimp alpha-glucosidase cDNA shows a high level of identity with that of the human sucrase-isomaltase, human maltase-glucoamylase, and human acid lysosomal alpha-glucosidase, indicating that the protein shares the same structural domains. The similarities among these proteins are found as clusters and characterize the glycosyl hydrolase family 31. To our knowledge, this is the first report to describe a satellite sequence in the 5′ genomic structure before the TATA box in an invertebrate sequence.

Key words. Alpha-glucosidase; cDNA nucleotide sequence; invertebrate; Crustacea.

Dietary starches are a mixture of two structurally different polysaccharides, amylose and amylopectin, which give rise to a branched configuration. Alpha-amylase (EC 3.2.1.1) is the endoenzyme that produces linear maltose oligosaccharides by hydrolysis of the alpha-1-4 linkage of amylose and branched alpha-dextrins or branched isomaltose oligosaccharides [1]. In mammals, these starch-derived oligosaccharides are hydrolyzed by small intestinal mucosal brush border-anchored sucrase-isomaltase (EC 3.2.1.48 and 3.2.1.10) and maltase-glucoamylase (EC 3.2.1.20 and 3.2.1.3) complexes [2]. The enzyme substrate specificities of sucrase-isomaltase overlap with those of maltose-glucoamylase. The glycoside hydrolase family 31, to which sucrase-isomaltase and maltase-glucoamylase belong, includes a range of enzymes with large specificity differences [3]. Alpha-glucosidase (EC 3.2.1.20) is one of the glycosyl hydrolases which plays a role in the final steps of starch digestion [4]. This enzyme is a member of a group of enzymes whose specificity is directed mainly towards the exohydrolysis of the alpha-1-4 glucosidic linkage and it rapidly hydrolyses oligosaccharides, in contrast to polysaccharides, which are hydrolyzed relatively slowly or not at all.

In general, alpha-glucosidases have a weak specificity and a given substrate is not strictly connected to a single type of protein. Alpha-glucosidases have been purified from microorganisms, plants, and animals, and several complete primary structures are known [5–10]. These glucosidases can be distinguished by their location, which can be lysosomal [6, 10, 11], anchored in the intestinal brush border membrane [12], or secreted in microorganisms [5].

* Corresponding author.

Alpha-glucosidase has been purified to homogeneity from the digestive gland of *Litopenaeus vannamei* [13]. In contrast to other known glucosidases, no alpha-1-6 glucose link hydrolysis activity has been observed in the shrimp purified enzyme [6, 11]. To gain information on the structure of the *L. vannamei* alpha-glucosidase, the entire sequence of this enzyme was deduced from the cDNA sequence. The primary structure and enzymatic substrate specificity were compared to evaluate the variation associated with the binding domains of this shrimp enzyme.

Materials and methods

Complementary DNA library. Messenger RNA from *L. vannamei* digestive gland was extracted using the RNA isolation kit from Stratagene. A lambda ZAP II cDNA library was established following the manufacturer's protocol (stratagene) [14]. This protocol produced cDNAs that were then inserted in the polylinker region of the pBluescript plasmid between *Eco*RI and *Xho*I restriction sites. Five million independent phages were obtained and after amplification, the library titer was 10^{10} pfu/ml.

PCR procedure. Two oligonucleotides were synthesized (Cybergene, France). The first was based on the partial amino acid sequence of the purified protein [13]. The second was designed as deduced from the amino acid sequence which is highly conserved in the catalytic site of other glucosidases. These oligonucleotides were used with oligo-dT-primed cDNAs synthesized from shrimp digestive gland total RNA for PCR amplification at 56 °C (Expand Taq polymerase from Boehringer).

The amplicons were analyzed on 1.5% agarose gels with *Pst*I-digested lambda DNA as size markers, extracted, and purified using the QIAquick gel extraction kit (Qiagen), ligated into the pGEM T-easy vector (Promega), and transformed into *Escherichia coli* JM 109 competent cells (Promega). The oligonucleotides used in the different experiments shown in this paper had the following sequences:

1)	5′ TGGMTNGAYATGAAYGAR 3′	(1357–1374)
2R)	5′ GRTTYCTNSWRAANGGRTARAA 3′	(1954–1933)
7R)	5′ CGGGTCGCTTGTTGGGGAACA 3′	(1720–1699)
9R)	5′ GAAGTTGGCCGGTTGGTTCAT 3′	(1440–1420)
WR1)	5′ GAAGCGACGTCCTGGCCCTGTGGA 3′	(126–98)
WR2)	5′ ACTCGATCGCCTGCAGCTTCGTCTGTTG 3′	(91–64)
pT3)	5′ GAAATTAACCCTCACTAAAGGG 3′	

QIA preps plasmid isolations (Qiagen) were carried out and the DNA was subsequently sequenced.

cDNA library screening procedure. The library was first screened using a 150-bp *Eco*RI/*Sac*II-restricted fragment of the cloned 300-bp amplimere as probe. Plaques were transferred to Hybond N+ membrane (Amersham) and screened with the 150-bp amplified cloned probe labelled with [^{32}P]dATP, using the random priming kit from Biolabs. Prehybridization of the duplicate membranes was achieved in a 50% formamide solution containing 1% SDS, 1 M NaCl, and sonicated salmon sperm DNA for 4 h at 42 °C. For hybridization, the denatured probe was added (10^6 cpm/ml) overnight at the same temperature. The filters were washed twice in 2 × SSPE, 0.1% SDS for 10 min at room temperature, then twice in 1 × SSPE containing 1% SDS at 65 °C for 15 min and then twice in 0.1 × SSPE for 5 min at 60 °C. The membranes were autoradiographed using Hyperfilm-MP (Amersham) with an enhancer screen. Positive clones which gave the strongest signal on plaque hybridization were isolated and characterized by double digestion with *Eco*RI and *Xho*I. The largest clones were selected for sequencing.

PCR on the digestive gland library. To access the complete sequence of the 5′ region of the cDNA, successive PCR experiments using a vector targeting primer (pT3) and reverse specific primers deduced from the selected cloned sequence were carried out directly on the digestive library extracted total DNA. PCR products were cloned in pGEM-T easy for further sequencing.

Gene walking. To determine the 5′ end of the alpha-glucosidase coding sequences, two specific 27-nucleotide-long antisense primers designed from the known 5′ sequence were used in successive rounds of amplification using adaptor-ligated genomic fragments as template. Genomic DNA was first digested with five different blunt-end restriction enzymes. Fragments obtained for each of these libraries were ligated to genome walker adaptors long enough to allow hybridization of two independent adaptor primers. Then a first PCR experiment was performed using two external primers [adaptor primer 1 and the specific antisense primer 1 (WR1)]. The PCR products obtained were diluted 50 times and 1 µl of the dilution was used for a second nested PCR experiment using the internal adaptor primer 2 and the internal specific primer 2 (WR2). These two PCR runs were performed with a two-temperature program: 94 °C, 25 s; 72 °C for 3 min and 7

cycles, then 94 °C, 25 s; 67 °C for 3 min and 32 cycles. The fragments obtained were cloned in pGEM-T easy and sequenced.

Sequencing. Sequencing was carried out by the dideoxy chain termination method using [^{35}S]dATP and sequenase on double-stranded DNA. All DNA sequences were determined on both strands using the T7 sequenase quick-denature plasmid sequencing kit (Amersham).

Results

PCR amplification. The first forward primer was designed from the conservative catalytic-site sequence WIDMNE (5′TGGMTNGAYATGAAYGAR3′) of alpha-glucosidase. The second oligonucleotide (reverse primer) was composed of 22 bases (5′GRTTYCTNSWRAANGGRTARAA 3′) based on part of the amino acid sequence which we obtained previously (HFYPFSRNH) [13].

After 35 cycles of amplification, a strong DNA band of about 550 bp was specifically obtained for the shrimp digestive gland RNAs. After extraction, the DNA amplicon was cloned and sequenced. A 549-bp-long sequence was determined and appeared to code for 183 residues of the central domain of the shrimp alpha-glucosidase. Alignment of the amplicon nucleotide sequence confirmed strong homology with various glycosyl hydrolases. The 549-bp-long amplified cloned product was digested with *Sac*II and *Eco*RI and the 150-bp *Sac*II/*Eco*RI-digested insert was used as probe to screen the *L. vannamei* digestive gland cDNA library.

cDNA library screening and deduced amino acid sequence. With the highly stringent conditions used, only seven clones (from a total of 15×10^4) were found to be positive. All appeared to contain an insert ranging between 600 and 1400 bp in length. The longest was named SAG (shrimp alpha-glucosidase). Analysis of this clone also confirmed the sequence of the probe and provided more information concerning the 3′ end of the messenger. Two successive PCR experiments on the library phage DNA using the PT3 primer permitted recovery of the nearly complete cDNA coding sequence with the exception of the first two residues at the 5′ end. Figure 1 illustrates the entire strategy which can be summarized as follows: (i) probe construction by PCR using degenerate primers on total digestive gland RNAs; (ii) screening of the digestive gland cDNA library by the probe (isolation of the entire 3′ region); (iii) successive PCR on the digestive gland cDNA library for more information on the 5′ region; (iv) gene walking.

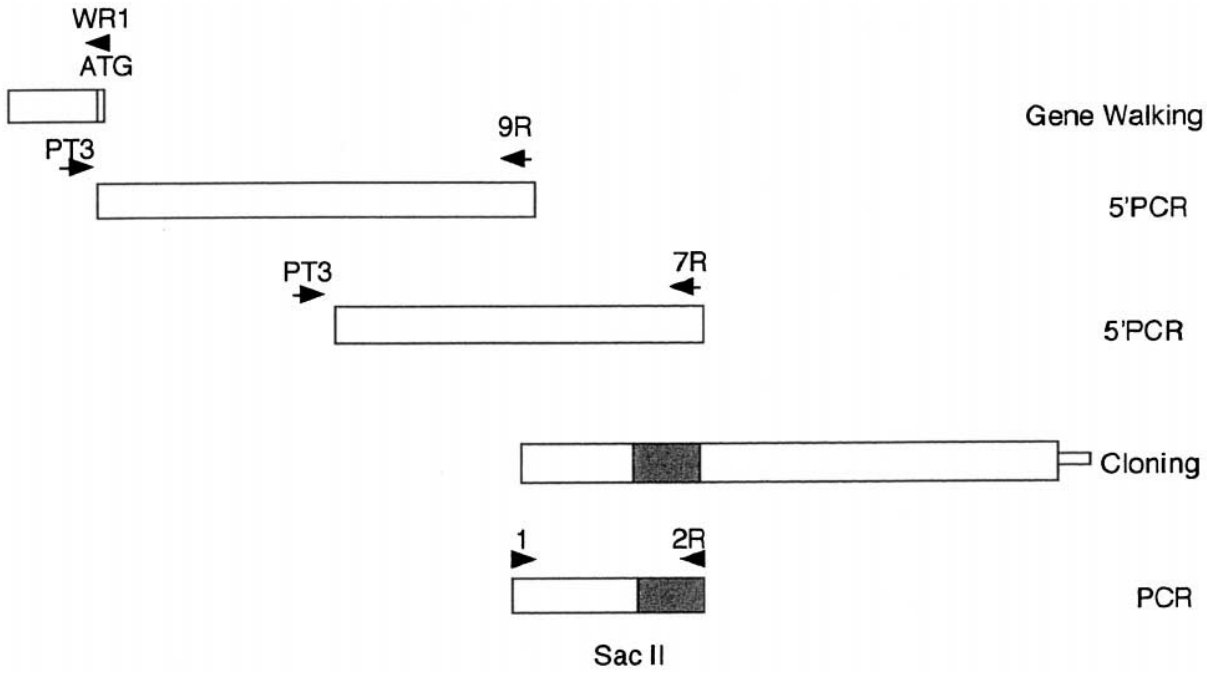

Figure 1. PCR and cloning strategies used to obtain the entire coding sequence of shrimp alpha-glucosidase. The probe used to screen the shrimp digestive gland cDNA library, was the 150-bp-long Sac II-digested fragment from the amplimere and is shown as a shaded box. Black arrows indicate the different primers used for the successive PCR steps.

```
gtcagaacgttcctcgggaaatataaccaaagaacttgaattcgtttctgtgagccacctggtgtaacagac  -106
agcaggtgtgcagtgcctgttggcgagggaaaagataatcaaaactcgacccggatataaggtggagacgcc   -34
                                        M  V  R  V  V  A  Q  V  I            -13
tccctccgaccccgtgttagctcgaagttgaag ATG GTG AGG GTG GTG GCG CAG GTT ATC        27
 L  A  V  A  L  T  G  S  L  S  L  G  Q  Q  T  K  L  Q                       6
CTG GCG GTG GCG CTG ACG GGG AGC CTT TCC CTC GGC CAA CAG ACG AAG CTG CAG      81
 A  I  E  C  P  F  P  E  G  Q  D  V  A  S  E  A  K  C                      24
GCG ATC GAG TGT CCA TTT CCA GAG GGC CAG GAC GTC GCT TCA GAG GCA AAG TGT     135
 A  E  Y  T  A  C  E  W  T  D  G  R  C  H  M  V  S  N                      42
GCT GAG TAC ACG GCC TGT GAA TGG ACG GAT GGA AGG TGT CAC ATG GTC AGC AAT     189
 D  V  G  G  Y  V  V  D  G  E  A  Q  D  T  G  R  G  F                      60
GAT GTG GGT GGC TAT GTG GTG GAC GGT GAG GCT CAG GAC ACC GGC AGG GGA TTC     243
 K  V  N  L  K  K  A  D  A  S  K  T  M  F  G  A  D  V                      78
AAG GTG AAC TTG AAG AAA GCA GAC GCC TCC AAA ACC ATG TTT GGT GCA GAT GTG     297
 A  D  L  V  F  E  V  I  Q  H  E  N  Y  H  V  Q  I  K                      96
GCC GAC CTC GTG TTT GAA GTT ATC CAG CAT GAG AAC TAC CAT GTG CAA ATC AAG     351
 I  Y  D  P  V  N  A  R  Y  E  V  P  L  P  L  N  L  P                     114
ATT TAC GAC CCG GTC AAC GCG CGC TAT GAG GTC CCG CTG CCA CTC AAC CTG CCA     405
 A  E  A  E  A  D  P  L  Y  S  V  S  V  S  G  N  G  E                     132
GCA GAG GCC GAG GCG GAT CCC CTG TAC TCC GTG AGC GTC AGC GGC AAC GGG GAG     459
 P  F  H  F  S  V  N  R  N  T  N  G  N  T  L  F  R  S                     150
CCC TTC CAC TTC AGC GTC AAC AGG AAC ACC AAC GGC AAC ACC CTT TTC CGA TCG     513
 E  G  P  L  T  F  E  D  Q  F  I  Q  L  H  T  G  L  M                     168
GAG GGA CCC CTG ACG TTT GAG GAC CAG TTC ATC CAA CTC CAC ACG GGA CTC ATG     567
 S  S  Y  L  Y  G  F  G  E  N  T  H  T  S  F  R  Q  V                     186
TCA TCC TAC CTG TAC GGT TTC GGC GAG AAC ACC CAC ACG TCT TTC CGC CAG GTG     621
 F  E  P  R  T  T  F  P  I  F  A  R  D  Q  P  V  G  T                     204
TTC GAG CCC CGC ACG ACC TTC CCA ATC TTC GCT AGG GAC CAG CCT GTG GGG ACG     675
 E  P  M  N  E  Y  G  H  H  P  Y  Y  M  V  M  E  D  D                     222
GAA CCT ATG AAC GAG TAC GGT CAC CAC CCT TAC TAC ATG GTC ATG GAG GAC GAC     729
 F  G  N  S  H  S  V  L  L  H  N  S  N  A  M  E  Y  S                     240
TTT GGG AAC TCT CAC TCC GTT CTT CTC CAC AAC AGC AAC GCC ATG GAG TAC TCC     783
 T  F  L  L  D  D  G  T  P  T  L  T  L  R  T  I  G  G                     258
ACC TTC CTT CTG GAC GAC GGG ACT CCC ACC CTC ACC CTG CGC ACC ATC GGC GGC     837
 I  I  D  L  H  F  F  L  G  P  D  P  E  D  L  N  L  Q                     276
ATC ATC GAT CTC CAC TTC TTT TTG GGT CCT GAT CCT GAA GAT TTG AAC CTT CAG     891
 Y  T  N  M  A  G  T  P  A  M  P  T  Y  W  S  L  G  F                     294
TAT ACA AAT ATG GCT GGC ACA CCC GCT ATG CCC ACG TAC TGG TCG CTG GGT TTC     945
 H  L  S  R  W  G  Y  N  S  T  D  G  V  R  A  A  R  E                     312
CAT CTT TCC CGC TGG GGT TAC AAC TCA ACA GAT GGC GTT CGT GCT GCC AGA GAG     999
 R  M  K  V  M  G  I  P  Q  D  V  Q  T  C  D  I  D  Y                     330
AGG ATG AAA GTT ATG GGC ATC CCT CAG GAT GTG CAG ACG TGT GAC ATC GAC TAC    1053
 M  D  R  Q  R  D  F  T  Y  D  P  V  S  W  G  D  M  P                     348
ATG GAC CGT CAA CGT GAT TTC ACT TAT GAT CCT GTC AGC TGG GGT GAC ATG CCA    1109
 D  L  I  N  E  L  H  N  D  N  I  K  V  T  L  I  L  D                     366
GAC CTC ATC AAT GAA CTG CAC AAT GAC AAC ATT AAG GTC ACA CTC ATT CTG GAT    1161
 P  A  L  V  I  D  F  D  N  Y  A  P  A  A  R  G  K  D                     384
CCT GCG CTG GTC ATC GAC TTC GAC AAC TAC GCC CCG GCA GCT CGA GGA AAG GAT    1215
 S  D  V  F  I  K  W  S  D  P  S  F  V  P  S  D  Q  E                     402
TCG GAC GTG TTC ATC AAG TGG TCC GAC CCG AGC TTC GTC CCG AGC GAT CAG GAG    1269
 A  G  T  D  D  Y  M  V  G  Y  V  W  P  D  T  K  T  I                     420
GCC GGA ACA GAT GAC TAC ATG GTG GGT TAC GTT TGG CCC GAC ACA AAG ACC ATT    1323
 F  P  D  F  L  K  P  E  T  A  T  W  W  T  N  E  L  K                     438
TTC CCC GAT TTC CTG AAG CCC GAG ACC GCA ACC TGG TGG ACC AAC GAG CTC AAG    1377
 L  F  H  D  A  L  Q  Y  D  A  I  W  I  D  M  N  E  P                     456
CTT TTC CAT GAT GCC CTG CAA TAC GAC GCC ATC TGG ATC GAC ATG AAC GAA CCG    1431
 A  N  F  G  T  N  L  D  K  P  W  N  W  P  E  E  R  T                     474
GCC AAC TTC GGC ACC AAC TTG GAC AAG CCT TGG AAC TGG CCC GAA GAA AGG ACC    1485
 P  G  V  K  C  P  E  N  K  W  D  S  P  P  Y  P  T  M                     492
CCT GGA GTT AAG TGT CCC GAG AAC AAG TGG GAC TCT CCC CCC TAC CCG ACC ATG    1539
 M  I  R  V  G  D  N  Q  S  K  K  I  S  D  H  T  I  C                     510
ATG ATC AGA GTT GGG GAT AAT CAG AGC AAG AAG ATT AGT GAT CAC ACC ATC TGC    1593
 M  S  G  N  O  T  D  G  T  K  T  Y  L  H  Y  D  V  H                     528
```

Fig. 2.

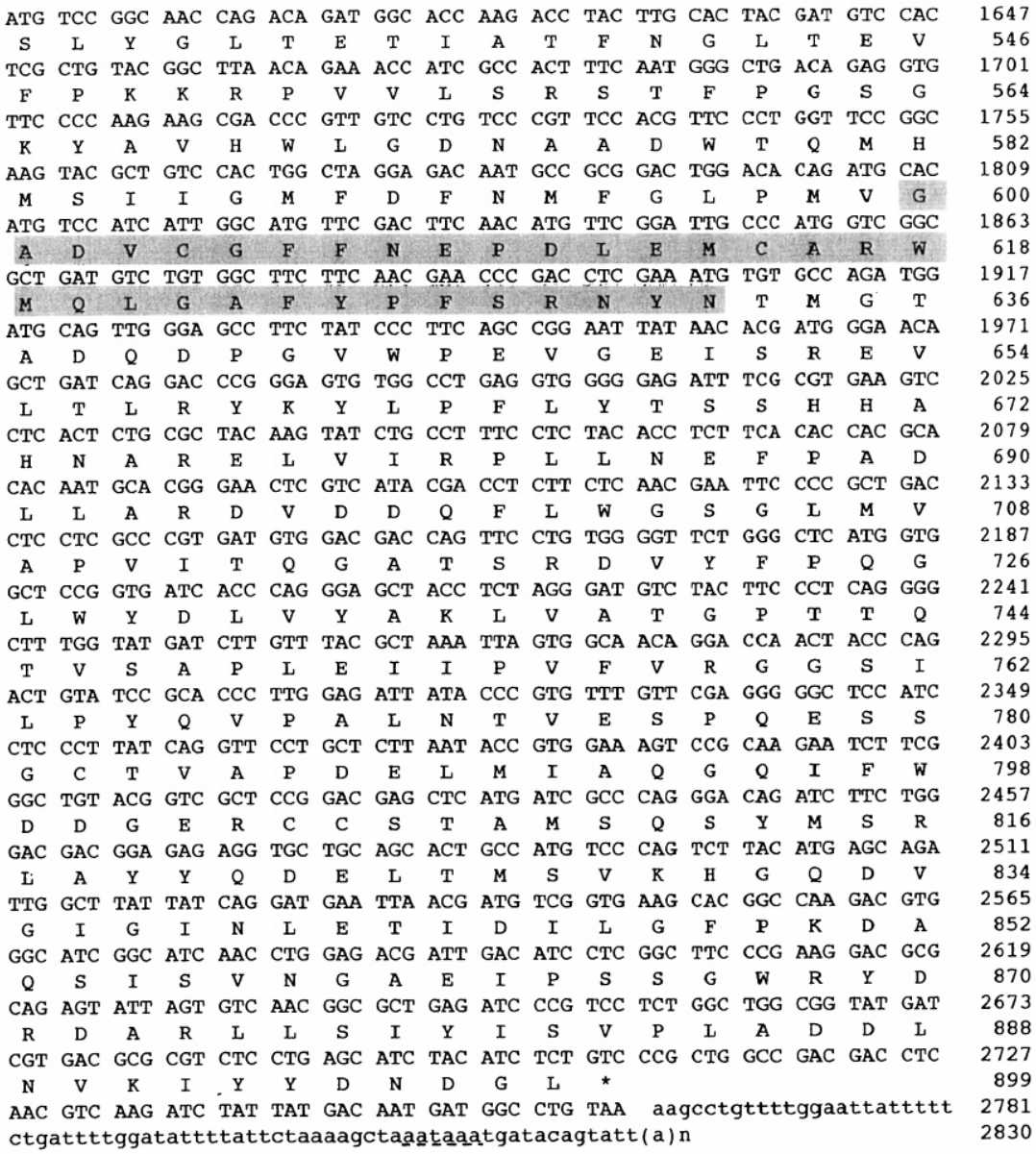

Figure 2. Nucleotide sequence of *Litopenaeus vannamei* alpha-glucosidase cDNA and its deduced amino acid sequence. The shrimp alpha-glucosidase peptide sequence is 919 amino acids long and has a molecular mass of 103,134 Da. The putative signal peptide corresponds to the first 20 amino acids (underlined). The putative catalytic amino acid sequence is underlined and the two glycosyl hydrolase family 31 signature sequences are gray-boxed. Shrimp glucosidase has three potential N-glycosylation sites (bold characters). Nucleotide numbering starts with the adenosine of the start codon. Amino acid numbering starts with the first N-terminal residue of the mature protein. The 3′ polyadenylation signal is indicated with a dashed line. The stop codon (TAA) is indicated by an asterisk.

The nucleotide sequence of the shrimp alpha-glucosidase has been deposited under GenBank accession number AJ250828 and is 2830 nucleotides long. The amino acid sequence deduced from the open reading frame from position 1 to position 2760 is shown in figure 2.

The SAG cDNA contains a 3′ noncoding sequence of 74 nucleotides. It also includes a polyA tract preceded (12 nt) by a consensus polyadenylation signal (AATAAA) at position 2812, suggesting that we have cloned the entire 3′ ultranslated region (UTR) of the mRNA.

Characterization of the putative promoter region. Only faint DNA bands in two among the five lanes could be seen after electrophoresis of the products of the first PCR step. After the second nested PCR step, bands of variable length were seen for all five libraries. The longest bands were two closely related 800-bp fragments corresponding to the *Dra*I digest, and two 1500-bp DNA products for the *Stu*I digest. Sequencing of the *Stu*I fragments allowed us to establish the coding sequence for the first residues of the preprotein and the promotor region of the gene (fig. 3). As expected, the *Dra*I site is found at a position 770 bp from the specific primer WR2 and the fragment ends with half of the *Stu*I site. The two closely related fragments show slight differences in nucleotides and size (around 50 nucleotides) that could be the expression of a length polymorphism of the satellite sequence. The repeated GA sequence is composed of $(GA)_3(GGGAGA)_{20}TCC(AG)_{34}AA(GA)_8$-$A(GA)_{14}$ and is located 372–128 nucleotides before the TATA box which starts at position -46.

Amino-acid homology. Comparative BLAST analysis of the shrimp alpha-glucosidase with human sucrase-iso-

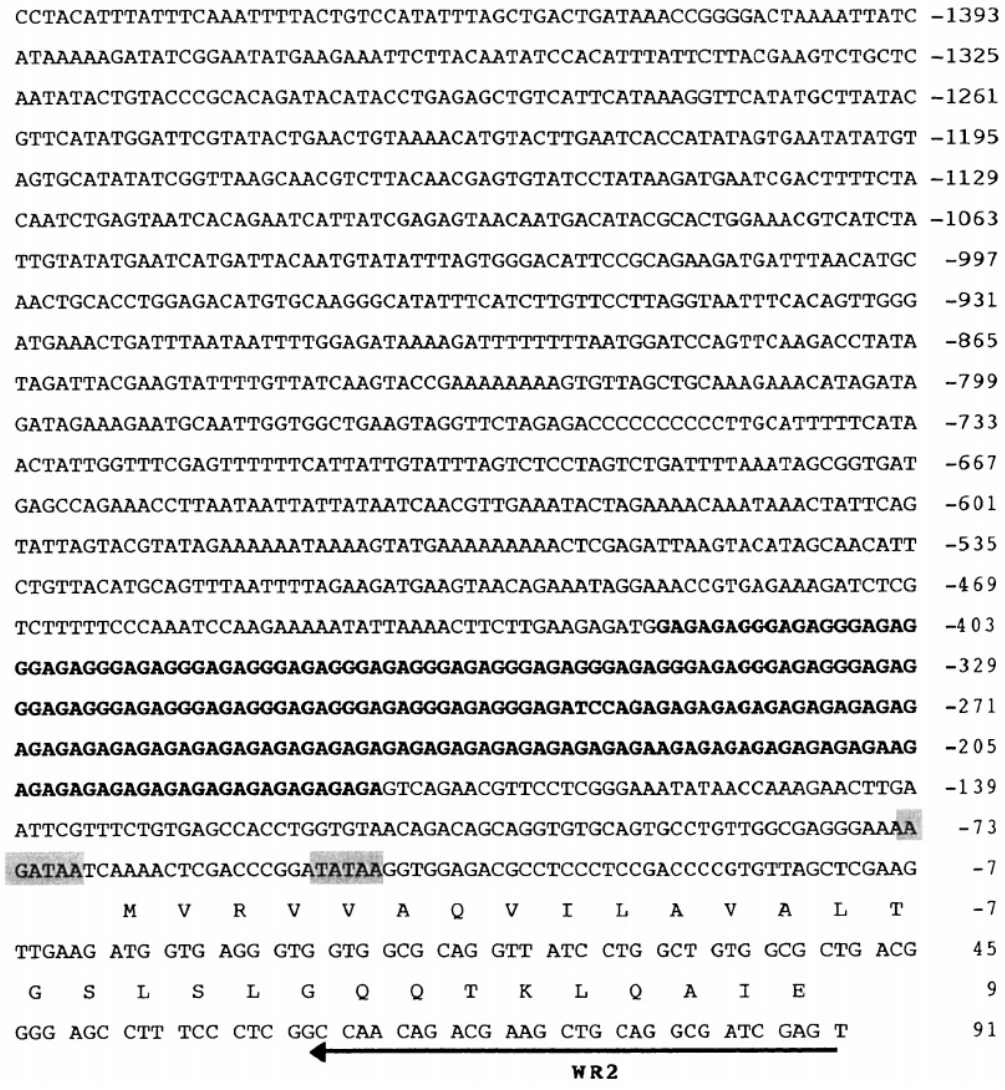

Figure 3. Nucleotide sequence of the 5′ region of the alpha-glucosidase gene as determined after a gene-walking experiment. Nucleotides corresponding to the WR2 primer hybridization site are underlined with a black arrow. Nucleotide numbering starts with the adenosine of the start codon. The repeated AG-rich sequence, a microsatellite, is shown in bold characters. The TATA and the GATA signal sequences are in gray boxes. Amino acid numbering starts with the first N-terminal residue of the putative mature protein.

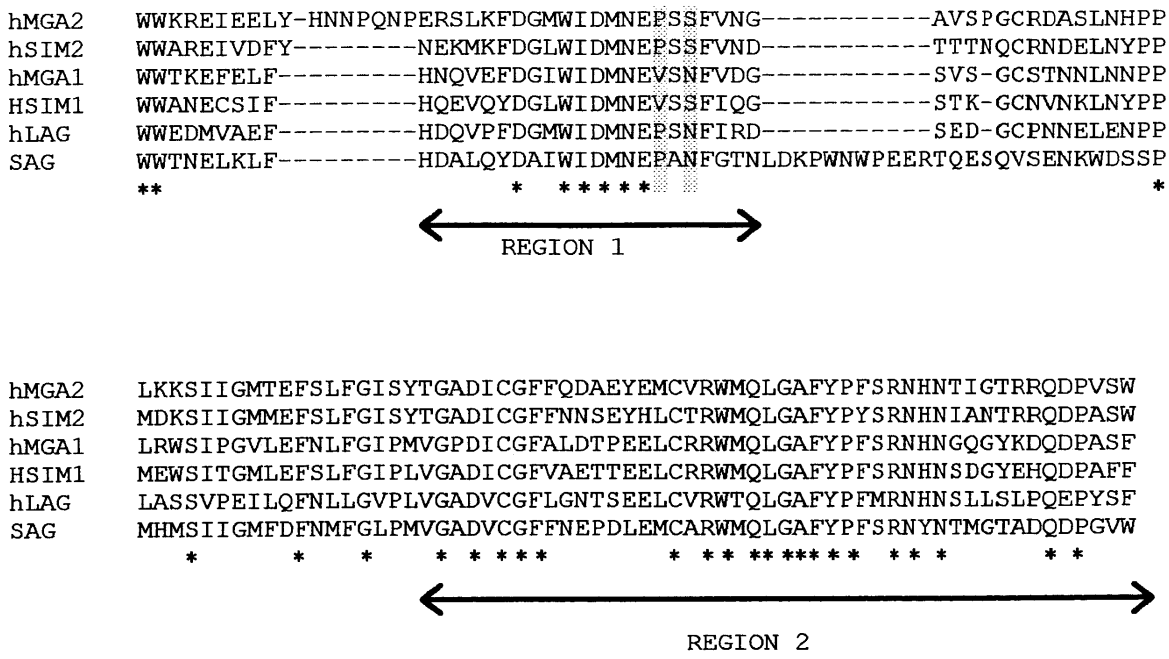

Figure 4. Alignments of the two putative catalytic amino acid domains (hMGA2), human glucoamylase 2; hSIM2, human sucrase-isomaltase 2; hMGA1, human glucoamylase 1; hSIM1, human sucrase isomaltase 1; hLAG, human lysosomal alpha-glucosidase; SAG, shrimp alpha-glucosidase [10, 17, 21]. Asterisks identify amino acids that are present at the same location in all six sequences. The two amino acid positions in region 1 that are implicated in the specificity of the catalytic activity are in a gray box.

maltase (hSIM), human maltase-glucoamylase (hM-GA), and human acid lysosomal alpha-glucosidase (hLAG) cDNA sequences reveals 38%, 38%, and 37% homology, respectively. Clusters of identical amino acids occurring throughout the sequence are revealed by amino acid sequence alignment of shrimp alpha-glucosidase with hSIM, hMGA and hAG sequences. The homologies around positions 470–475 and 621–653 are the most evident (fig. 4). In fact, these two domains are the signature of the glycosyl hydrolase family 31 to which these enzymes belong.

Discussion

Peptide sequence and homology. We isolated cDNA clones encoding shrimp alpha-glucosidase. The combined length of the clones is around 3 kbp containing nearly the entire coding sequence and the complete 5′ and 3′ UTRs. The coding region of the shrimp alpha-glucosidase cDNA reveals a protein of 919 amino acids (without the first methione) with a calculated molecular mass of 103,134 Da. This result confirms the molecular mass estimation by SDS PAGE with the purified protein [13]. A putative signal peptide corresponds to the first first 21 amino acids. The use of the Prot Param Tool program from the Exapsy Server shows that there is a most likely cleavage site between glycine and glutamine after the stretch of hydrophobic residues in the N-terminal sequence. These features of the amino acid sequence fulfil the requirement for a signal peptide common to lysosomal or secretory proteins [15, 16]. Alignment of the deduced amino acid sequence was established with the series of human lysosomal acid alpha-glucosidase, human sucrase-isomaltase and human maltase-glucoamylase using the program Clustal. Two typical signature sequences of the glycosyl hydrolase family 31 are shown in bold in figure 4. The first is the catalytic site which is strongly conserved in other carbohydrate hydrolases, including those of mammals, fungi, and plants. This type of catalytic site (WIDMNE) present in the shrimp alpha-glucosidase has an aspartic acid (D) catalytic site [10, 17, 18]. The second glycosyl hydrolase family 31 signature is rich in aromatic residues (fig. 4). This indicates that both motifs could constitute markers for genes derived from the same ancestor [19].

On the basis of sequence comparison, we cannot establish whether the shrimp alpha-glucosidase is a lysosomal or secreted form. But we think that the present sequence most resembles the acid alpha-gucosidase because no membrane-spanning region was identified, and no ap-

propriate sequence homologies were found with sucrase-isomaltase or maltase-glucoamylase.

5′ structure of the shrimp alpha-glucosidase gene. The 5′ UTR, of which around 1.5 kb has been sequenced, contains a typical Goldberg-Hogness sequence and a TATA box, but no CAAT box. Furthermore, before the TATA and the GATA signal sequences, a repeated AG-rich sequence is found. The presence of a microsatellite in the regulatory part of an invertebrate gene has not been described previously, although a microsatellite has been found in introns of digestive enzyme genes in shrimp [unpublished data]. Chantret et al. [19] reported a 5′ upstream region of the human sucrase-isomaltase gene, of which 6.3 kb have been sequenced: typical CAAT and TATA boxes were identified but no repeated sequence was found.

The gene-walking method provided information not only on the untranscribed genomic region but also for the 5′ end of the cDNA. We do not know the size of the 5′ UTR, but it resembles previously identified 5′ UTRs from different cDNAs encoding digestive enzymes in shrimp, we expect this region to be short (around 20 bp).

Glycosylation. Shrimp alpha-glucosidase has three potential N-glycosylation sites, and a total of five sites for potential O-glycosylation. How many of the potential sites are actually glycosylated is not known. Comparing the deduced size of the shrimp alpha-glucosidase from the cDNA sequence with that obtained by SDS-PAGE electrophoresis of the purified enzyme, the difference between the two molecular masses is small: either the mature shrimp alpha-glucosidase is not much glycosylated, or during the purification steps, the enzyme loses glycosyl components. In human acid lysosomal alpha-glucosidase, seven possible glycosylation sites were identified (Asn-X-Ser/Thr) [10], a number close to that found in shrimp alpha-glucosidase. In contrast, in mammalian glucoamylase-maltase and sucrase-isomaltase, carbohydrates make up about a third of the molecular mass of the mature enzyme [12, 20].

Substrate specificity. The shrimp alpha-glucosidase sequence and hSIM (which exhibits alpha-1-6 glucosidic link hydrolysis) differ most at the amino- and carboxy-terminal ends. These nonhomologous regions have a function in enzymatic specificity and may in part determine differences in substrate specificity [10, 21]. Computer-assisted comparison of these sequences reveals interesting similarities and differences.

Naim et al. [22] noted that the putative active site of the glycoside hydrolase family 31 (signature region 1) which is a stretch of ten amino acids reveals only two amino acid substitutions (VP and NS or vice versa, for positions 476 and 478, respectively) between alpha-glucosidase and isomaltase. They suggested that substituting V for P and N for S in the putative active site of these enzymes could be important in providing the corresponding enzyme with a conformation that has higher affinity for one type of dissaccharide (alpha-1-4 linkage) than for the other (alpha-1-6 linkage). Thus, the absence of alpha-1-6 glucose link hydrolysis activity by the shrimp alpha-glucosidase could not be explained on the basis of a primary struture comparison alone.

Frandsen and Svensson [23] listed 24 sequences of family 31 glycosyl hydrolases and proposed an alignment that covers these two consensus sequences (referred to signature regions 1 and 2). Five residues in region 1 and nine in region 2 are invariant. They supposed that these two signature regions contain several potential substrate-binding residues. To resolve structure/function relationships in the glycoside hydrolase family 31, they believe that three-dimensional structure determination coupled with mutational analysis are needed to understand the mechanism of action and specificities represented in this type of enzyme. But we still await the first three-dimensional structure for a member of family 31 in animals. Meanwhile, further biochemical characterization will help us to understand the relationship between structure and function in shrimp alpha-glucosidase, and the cDNA will be a useful tool to determine which cell type is producing alpha-glucosidase in the digestive gland, and to study at the molecular level (mRNA expression, for example) the induction of alpha-glucosidase activity by different dietary carbohydrates [24].

1 Hassid W. Z. and Ball C. E. (1957) The carbohydrates. In: Chemistry, Biochemistry, Physiology, pp. 501–506, Pigman W. (ed.), Academic Press, New York
2 Fogel M. R. and Gray M. G. (1973) Starch hydrolysis in man: an intraluminal process not requiring membrane digestion. J. Appl. Physiol. **35:** 263–267
3 Henrissat B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. **280:** 309–316
4 Gray M. G. (1992) Starch digestion and absorption in nonruminants. J. Nutr. **122:** 172–177
5 Shanta Kumara M. M. C., Decort S. and Verachtert M. (1993) Localization and characterization of alpha-glucosidase activity in *Brettanomyces lambicus*. Appl. Environ. Microb. **63:** 2352–2358
6 Bruni C. B., Autrichio F. and Covelli I. (1969) Acid alpha-glucosidase glucohydrolase from cattle liver. J. Biol. Chem. **244:** 4735–4742
7 Killilea S. D. and Clancy M. J. (1978) Properties of potato alpha-glucosidase. Phytochemistry **17:** 1429–1431
8 Kinsella B. T., Hogan S., Larkin A. and Cantwell B. A. (1991) Primary structure and processing of the *Candida tsukubaensis* alpha-glucosidase, homology with the rabbit intestinal sucrase-isomaltase complex and human lysosomal alpha-glucosidase. Eur. J. Biochem. **202:** 657–664
9 Sugimoto M., Furui S. and Suzuki Y. (1997) Molecular cloning and characterization of a cDNA encoding alpha-glucosidase from spinach seeds. Plant. Mol. Biol. **33:** 765–768
10 Hoefsloot L. E., Hoogeveen-Westerveld M., Kroos M. A., Van Beeumen J., Reuser A. J. J. and Oostra B. A. (1988)

Primary structure and processing of lysosomal alpha-glucosidase: homology with the intestinal sucrase-isomaltase complex. EMBO J. **7:** 1697–1704
11 Jeffray P. L., Brown D. H. and Brown B. I. (1970) Studies of lysosomal alpha-glucosidase: purification and properties of the rat liver enzyme. Biochemistry **9:** 1403–1415
12 Naim H. Y., Sterchi E. E. and Lentze M. J. (1988) Structure, biosynthesis and glycosylation of human small intestinal maltase-glucoamylase. J. Biol. Chem. **263:** 19709–19717
13 Le Chevalier P. and Van Wormhoudt A. (1998) Alpha-glucosidase from the hepatopancreas of the shrimp, *Penaeus vannamei*. J. Exp. Zool. **280:** 384–394
14 Sellos D. and Van Wormhoudt A. (1992) Molecular cloning of a cDNA that encodes a serine protease with chymotrypsic and collagenolytic activities in the hepatopancreas of the shrimp *Penaeus vannamei* (Crustacea Decapoda). FEBS Lett. **309:** 219–224
15 Watson M. E. E. (1984) Compilation of published signal sequences. Nucleic Acids Res. **12:** 5145–5164
16 Nielsen H., Engelbrecht J., Brunak S. and Von Heijne G. (1997) Identification of prokaryotik and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. **10:** 1–6
17 Hunziker W., Spiess M., Semenza G. and Lodish H. F. (1986) The sucrase-isomaltase complex, primary structure, membrane-orientation, and evolution of a stalked, intrinsic brush border protein. Cell **46:** 227–234
18 Hermans M. M. P., Kroos M. A., Van Beeumens J., Oostra B. A. and Reuser A. J. J. (1991) Human lysosomal alpha-glucosidase, characterization of the catalytic site. J. Biol. Chem. **266:** 13507–13512
19 Chantret I., Lacasa M., Chevalier G., Islam I., Mantrei N., Edwards Y. et al. (1992) Sequence of the complete cDNA and the 5′ structure of the human sucrase-isomaltase gene. Biochem. J. **285:** 915–923
20 Rodriguez I. R., Taravel F. R. and Whelan W. J. (1984) Characterization and function of pig intestinal sucrase-isomaltase and its separate subunits. Eur. J. Biochem. **143:** 575–582
21 Nichols B. L., Eldering J., Avery S., Hahn D., Quaroni A. and Sterchi E. E. (1998) Human small intestinal maltase-glucoamylase cDNA cloning. J. Biol. Chem. **273:** 3076–3081
22 Naim H. Y., Niermann T., Kleinhans U., Hollenberg C. P. and Strasser A. W. (1991) Striking structural and functional similarities suggest that intestinal sucrase-isomaltase, human lysosomal alpha-glucosidase and *Schwanniomyces occidentalis* glucoamylase are derived from a common ancestral gene. FEBS Lett. **294:** 109–112
23 Frandsen T. P. and Svensson B. (1998) Plant alpha-glucosidases of the glycoside hydrolase family 31: molecular properties, substrate specificity, reaction mechanism, and comparison with family members of different origin. Plant Mol. Biol. **37:** 1–13
24 Le Chevalier P., Sellos D., Le Boulay C., Lecher P. and Van Wormhoudt A. (1999) Alpha-glucosidase in Crustacea: cDNA characterization (abstract). In: Gene Regulations by Nutrients and Hormones, 26th Meeting of the Federation of European Biochemical Societes, Nice, 19–24 June, p. 296.