Biologia, Bratislava, **57/Suppl. 11:** 191—196, 2002

Structure of the of α-amylase genes in crustaceans and molluscs: evolution of the exon/intron organization

Daniel Y. Sellos* & Alain Van Wormhoudt

Station de Biologie Marine du Museum National d'Histoire Naturelle, BP 225, F-29900 Concarneau, France; tel.: ++ 33 2 98 97 06 59; fax: ++ 33 2 98 97 81 24; e mail: sellos@mnhn.fr

Sellos, D. Y. & Van Wormhoudt, A., Structure of the of α-amylase genes in crustaceans and molluscs: evolution of the exon/intron organization. Biologia, Bratislava, **57/Suppl. 11:** 191—196, 2002; ISSN 0006-3088.

α-Amylase is present in all groups of animals and characterized by an eight-stranded β/α-barrel three-dimensional structure. The exon-intron organization of the gene is highly variable with respect to the number of introns and the location of the junctions. Phylogeny, inferred from the structure of the gene, cDNA or amino-acid sequences including information from marine invertebrates, revealed interesting results especially concerning the proposed position of *Drosophila*, particularly since the structure of their genes reveals one or no intron. Different hypothesis could be suggested, such as the disappearance of introns during evolution of the insect group with specific and (or) rapid adaptation to new substrates. Intron positions were determined in crustaceans and in molluscs. Among Arthropoda, Crustacean penaeids are very ancient, as they were already present in the Devonian period. The complex structure of their α-amylase genes revealed the presence of nine introns. The position and phase of some of these introns are also seen in insects. Two are also present in *Ceratitis*. One or no intron is usually observed in the Drosophila family. In the bivalve mollusc, *Crassostrea gigas*, we have determined that the gene contains seven introns. Changes in the location of intron insertion are reported mainly in the 5' end of the genes and are discussed. Many introns from one gene having no counterpart in the others, seemed to have appeared later in evolution at a non-specific position at the end of the gene. More observations however are necessary to confirm these hypotheses, taking into account other members of the α-amylase family from more primitive eukaryotes in different groups such as poriferes, cnidaires, ecdysozoa and lophotrochozoa.

Key words: amylase, gene, evolution, adaptation, invertebrates, marine, populations.

Introduction

α-Amylase and α-amylase precursors were extensively studied to acquire fundamental knowledge, but also for interest in breeding programs. The sum of characterized sequences from prokaryotes, mammals and plants allow us to use this information as markers for evolution, adaptation processes, phylogeny or population studies (Janecek, 1994, 1997). The choice of this func-

* Corresponding author

tional marker to characterize populations has multiple interests as it could be used as a genetic marker, and it could give information on the capabilities of carbohydrate assimilation of individuals (Inomata et al., 1995; Moal et al., 2000). Activities of digestive enzymes constitute a physiological parameter affecting the digestive capacity (Ibarrola et al., 2000); the control of such a parameter may be important to maximize the assimilation rate and consequently the level of energy gain. The shrimp *Penaeus vannamei* is one of the most cultivated crustacean species in the world and the oyster *Crassostrea gigas* represents the first production of the French conchyliculture. It was recently introduced in France following large-scale destruction of endemic species (Grisel & Heral, 1991). The increase of production is correlated to a decrease of productivity related to a limiting trophic capacity of the environment. In this respect, the knowledge of species, populations and individual genetic potentials relating to specific adaptative capabilities is of importance. Variation of α-amylase isoforms has already been used for population studies in insects (Cariou & Da Lage, 1993; Dainou et al., 1993).

To date, knowledge of the sequence of α-amylase genes is inavailable for marine bivalve species and crustaceans and is limited for invertebrates except insects, where some introns were observed (Boer & Hickey, 1986; Grosmann & James, 1993; Da Lage et al., 1992, 1996, 2000). These authors have demonstrated the existence of differential expression and different chromosome localization for these genes. In molluscs, the structure of the α-amylase cDNA was determined in *Pecten maximus* (Le Moine et al., 1997); in the oyster, from a digestive gland cDNA library, only one cDNA was recovered although many different isoforms were determined by electrophoresis (Moal et al., 2000).

In this paper, we report the characterization of the structure of the tropical shrimp α-amylase genes and unlike the situation in insects, the existence of numerous introns (nine) was reported, as is found in vertebrates. We also report the structure of the α-amylase genes in the mollusc bivalve *Crassostrea gigas*, the level of expression of α-amylase isoforms in different tissues and the characterization of the level of polymorphism observed in different oyster populations sampled along the Atlantic French coast.

Material and methods

An unrooted neighbor-joining tree was constructed using mature α-amylase protein sequences. Crustacean and molluscan α-amylase sequences (accession numbers: AJ133526, AJ133379, AJ133119, AF320688 and AF321515 and X99729 respectively), vertebrate α-amylase sequences (*Homo sapiens*: M28443; *Sus crofa*: AF064742; *Mus musculus*: J00359; *Rattus rattus*: J00703 and *Gallus gallus*: U63411) and insect α-amylase sequences (*Aedes aegypti*: AF000569; *Anopheles merus*: U01210; *Culex tarsalis*: U01211; *Drosophila melanogaster*: X04569) were aligned with *Caenorhabditis elegans* α-amylase (T20090), *Streptomyces griseus* α-amylase (X57568) and *Alteromonas haloplanctis* α-amylase (X58627) using ClustalW (Thompson et al., 1994). Extensions in bacterial and some insect α-amylase sequences were removed to allow a correct alignment. Bootstrap values were calculated on 100 replicates.

A phylogenetic tree was constructed using uncorrected and corrected distances through the Neighbor-Joining approach using Protpars (Felsenstein, 1996).

Results and discussion

In the shrimp, a total number of eight α-amylase electromorphs were characterized by specific electrophoretic migration on agarose gel (data not shown). A high degree of polymorphism was observed in the number and type of isoforms from one individual to another inside the same population and between populations from different areas. To try to explain this high number of expressed isoforms, we investigated the structure of the α-amylase gene, with regards to the existence of possible gene variants and putative allelic polymorphism.

In crustaceans as in molluscs, the α-amylase genes have a complex structure with an interrupted coding sequence (Fig. 1). In the mollusc *Crassostrea gigas*, the two α-amylase genes are characterized by the same organization, with the existence of eight exons separated by seven introns. Sizes of exons vary from 37 bases for exon 8 to 521 bases for exon 4. When homologous exons from the two genes are compared, sizes are similar except for a deletion in gene B of the three nucleotides coding for the glutamine in position 279 in gene A. Sizes for introns are different inside a given gene (for example in gene A, from 123 to 1309 nucleotides), but also for a given intron in the two genes. A microsatellite repeat (TC_{37}) is located in intron 4 of gene A, starting three nucleotides after the GT donor site. It represents 47% of the whole intron. In gene B, the highest corresponding TC repeat contains only three consecutive repetitions of this motif. All the exon/intron junctions are found at the same locations and with

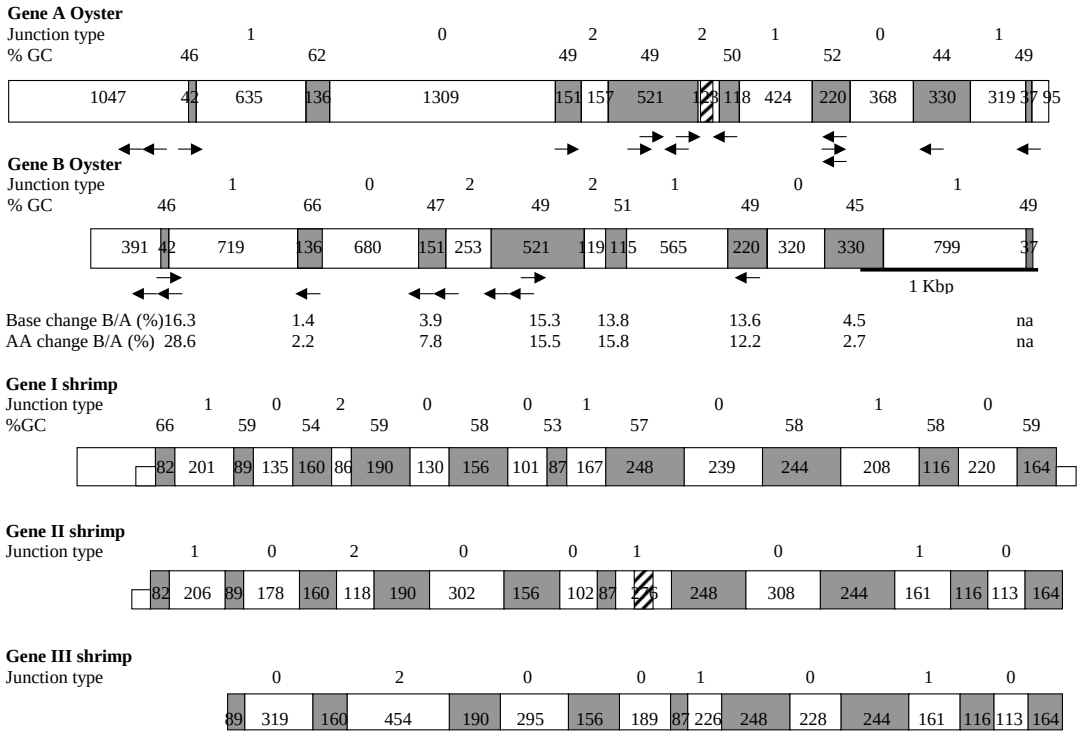

Fig. 1. Schematic representation of the α-amylase gene structure for the two molluscan and the three crustacean genes. The grey boxes represent the exons. Arrows indicate the oligonucleotide primers that were used to amplify specific introns, part of or the whole gene. Sizes of introns and exons, type of junction and GC percentage are indicated. Microsatellite sequence is indicated by a striped box.

a homologous type for the two genes. The two encoded proteins have also the same type of organization, with the presence of a 20-residue-long signal peptide and a 499-500-residue-long mature enzyme. For the whole coding sequence, 11% of the nucleotides are changed. The highest level of substitution is observed for exon 1 (16% of the nucleotides of gene B when compared to A), and for exons 4, 5 and 6. The lowest level of change is seen for exon 2 with only 1.6%. This exon has also the highest value of GC content with 62–66% when all the others have a GC level between 44 and 52%. The same relationship is not confirmed for exon 7 with a GC value close to the lowest (45%) and a few changes in nucleotides (4.5%). Overall, for the complete nucleotide sequences of the genes, around 60% identity was determined between A and B.

In the crustacean, several isoforms of the α-amylase gene were characterized and could be assembled in three sub-families (Fig. 1). All the genes showed the same complex organization with the existence of ten exons forming the coding sequence for the pre-α-amylase interrupted, with nine introns. The sizes of the introns are different in the same gene and different for the same intron in the different genes (from 86 bp for the smallest in gene "I" to 454 bp for the longest in gene "III"). The two last introns for genes II and III have the same length and sequence, showing that they may have come from a recombinant process. As in the case of the oyster, one of the shrimp genes (II) contains a microsatellite with a complex repeated sequence: "(CA)4(TA)2(CA)4(TATG)2(TG)2TTC(GT)11 GCTTG(TA)3TC(AT)3C(AT)10(TA) 34". The sizes of the exons in the shrimp genes are the same (sequences of genes II and III are partial, missing 5' and 3' portions). When we compare GC contents of the exons of the genes, it appears that crustacean α-amylase genes show equally-distributed high amounts of GC while only exon 2 from the mollusc is rich in GC. This should give a high stability to the crustacean genes. The few annotated structural gene sequences available for these two marine invertebrates in the databases do

not actually allow us to make useful comparisons.

If we compare the deduced amino acid sequences from the molluscan α-amylase genes, we find 54 residues are changed in protein B compared to A. Here too, the highest percentage difference is found for the part of the protein encoded by exon 1 (and coding for the first part of the signal peptide) with 29% changes; then for a large part of the protein from residues 91 to 378 (encoded by exons 4 to 6), there is also an appreciable change. The parts showing the lowest amino acid changes are encoded by exon 2 (2.1% of residues modified between AA-7 and AA91) and by exon 7 (2.7% of change between AA379 and the C-terminal residue). The same level of amino acid change (11.5%) was found between different shrimp alpha-amylase variants (Van Wormhoudt & Sellos, 1996). The overall identity of protein A is 90% with protein B, 72% with another bivalve mollusc α-amylase (Le Moigne et al., 1997) and 60% with human α-amylase.

The amounts of mRNA transcripts from the two molluscan genes, amplified with RT-PCR using RNA extracts from the different digestive tissues and specially designed primers (Ha-aH and Hb-bH) for targets specific for gene A or B, were compared to those of transcripts produced by the actin gene with specific actin primers AP1 and AP2 (Unger and Roesijadi, 1993). As shown in mammals, (Hagenbuchle et al., 1980), it appears that even if the two genes are expressed in the different digestive tissues, gene A is found to be strongly preferentially expressed in the digestive gland, at a lower level in the stomach, and at a low level in labial palps, mantle and intestine (data not shown). Gene B is preferentially expressed in the labial palps (to be compared to mammalian salivary α-amylase) and, at a lower level, in the other different tissues. In the tropical shrimp *Litopenaeus vannamei*, the three genes were found to be expressed only in the digestive gland (Le Moullac et al., 1996).

With respect to the insertion positions of introns inside the coding gene sequences (Fig. 2), it appeared all the eukaryotic α-amylase genes have an interrupted coding organization, the most complex being human and crustacean. For insects, *Drosophila*, with only one intron, has the simplest organization but this is not representative of the group, for *Apis* and *Ceratitis* have two introns (Ohashi et al., 1999; D. A. Hickey, unpublished results), *Tribolium*, three and *Ostrinia* six introns. For the different genes of any one species, all the exon/intron junctions are found at the same location with the same type of junction at the same amino acid. If we compare the different groups, we find the complexity of organization decreases in the order vertebrates (human), crustaceans, molluscs, worms and the insect *Ostrinia*. The first intron, that is present in molluscs and in crustaceans is located close to the sequence coding for the signal peptide/mature enzyme junction. The second intron is strongly conserved during evolution. The third intron is found at the same location in mollusc, shrimp and worm and is present, but shifted, in humans. The fourth intron is found in the crus-

Organism / Intron No.	*Litopenaeus vannamei*	*Drosophila ananassae*	*Drosophila melanogaster*	*Ceratitis capitata*	*Ostrinia*	***Tribolium***	*Apis*	*Caenorabditis elegans*	*Crassostrea gigas*	*Human*	*Limulus polyphem*
Intron 1	**Q0**								**V1 ***		
Intron 2	**Q0**	**Q0**		**Q0**	**Q0**	**Q0**			**Q0**	**Q0**	
Intron 3	**R2**							**R2**	**R2**	**G0 ***	
Intron 4	**Q0**				**R0**	**Q0 ***		**H0**		**Q0**	**Q0**
Intron 5	**K0**		**E0**	**E0**	**R2 ***		**R2 ***	**D0 ***			
Intron 6	**G1**					**G1**				**E0 ***	
Suppl. intron					**R2**				**R2**	**K2**	
Intron 7	**K0**						**K0**		**G1**	**R2 ***	
Suppl. intron										**N0**	
Intron 8	**G1**				**S0 ***			**A0 ***	**W0 ***	**R2 ***	
Intron 9	**Q0**				**Q0**			**K2 ***		**W2 ***	
Suppl. intron								**D0**	**G1**		

Fig. 2. Comparison of the position of the introns in different α-amylase genes. The junction is indicated by a letter representing the amino acid and a number giving the type of the junction (0 = the junction occurs after the concerned amino acid, 1 = the junction is found after the first nucleotide of the codon, and 2 = the junction is observed after the second nucleotide of the concerned codon). The junction phases found in corresponding positions in at least two species are boxed. An asterisk indicates a quasi-conserved position.

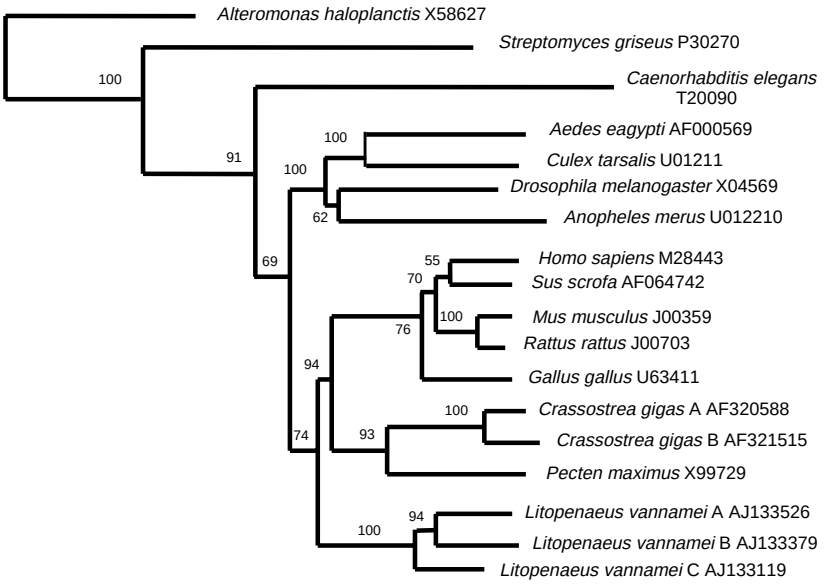

Fig. 3. Unrooted neighbor-joining tree (made with Clustal W) of the α-amylase coding sequences. The accession numbers and other details are given in the Material and methods section.

tacean, some insects, the worm, humans and in the primitive arthropod; *Limulus polyphemus* (D. Y. Sellos, unpublished results). Alignment of the location of the other junctions is more difficult to establish as introns could be absent or shifted by tens of nucleotides. It is interesting to note that the well-conserved locations of introns 2 and 4 in a large number of species are also found in the shrimp that is considered to be very ancient since it was already present in the Devonian period. In the primitive arthropod *Limulus polyphemus* we were able only to amplify a small part of the gene corresponding to a part of exon 4, intron 4 and a part of exon 5 (following the shrimp numbering; D. Y. Sellos, unpublished results). The fact that we found intron 4 in the same location with the same junction phase could argue in favour of the ancient presence of this intron and the possible disappearance of the intron during evolution in insects and other groups. After establishment of the alignment of the α-amylase protein sequences (based on 504 residues), the phylogenetic tree was constructed showing the relationship between groups, using 436 variable sites among which 331 positions are informative. The tree places crustaceans close to vertebrates and molluscs, and further from insects (Fig. 3). In this manner, arthropods appeared as a paraphyletic group. This tree organization is also shown by Da Lage et al. (2002), using a large number of insect sequences. This finding could be explained by the presence of several deletions in insect sequences, some being shared with prokaryotes. Even removing all the sequence portions that were not aligned because of the presence of deletions in one of the sequences, the same tree is obtained, always giving arthropods as a paraphyletic group. This may be due to the fact that insects, appearing recently, have rapidly diverged while penaeids evolved slowly during a long period of time, keeping ancestral features. Other marine invertebrate α-amylase sequences would help to solve the problem. We are currently working to determine α-amylase gene organization of primitive crustaceans (Polyplacophore) and the arthropod (*Limulus polyphemus*).

References

Boer, P. M. & Hickey, D.A. 1986. The α-amylase gene in *Drosophila melanogaster*: nucleotide sequence, gene structure and expression motifs. Nucleic Acids Res. **14:** 8399–8411.

Cariou, M. L. & Da Lage, J. L. 1993. Isozyme polymorphisms, pp. 160–171. In: Tobari, Y. N. (ed.) *Drosophila ananassae*, Genetical and Biological Aspects, Japan Scientific Societies Press, Tokyo.

Dainou, O., Cariou, M. L., Goux, J. M. & David, J. R. 1993. α-Amylase polymorphism in *Drosophila melanogaster*: haplotype frequencies in tropical African and American populations. Gene. Sel. Evol. **25:** 133–155.

Da Lage, J. L., Lemeunier, F., Cariou, M. L. & David, J. R. 1992. Multiple α-amylase genes in *Drosophila ananassae* and related species. Gen. Res. Camb. **59:** 85–92.

Da Lage, J. L, Wegnez, M. & Cariou, M. L. 1996. Distribution and evolution of introns in *Drosophila* α-amylase genes. J. Mol. Evol. **43:** 334–347.

Da Lage, J. L., Maczkowiak, F. & Cariou, M. L. 2000. Molecular characterization and evolution of the α-amylase multigene family of *Drosophila ananassae*. J. Mol. Evol. **51:** 391–403.

Da Lage, J. L., Van Wormhoudt, A. & Cariou, M. L. 2002. Diversity and evolution of the α-amylase genes in animals. Biologia, Bratislava **57 (Suppl. 11):** in press.

Felsenstein, J. 1996. Inferring phylogenies from protein sequences by parsimony, distances and likelihood methods. Methods Enzymol. **266:** 418–427.

Grizel, H. & Heral, M. 1991. Introduction into France of the Japanese oyster (*Crassostrea gigas*). J. Cons. Int. Explor. Mer **47:** 399–403.

Grosmann, G. L. & James, A. A. 1993. The salivary gland of the vector mosquito *Aedes aegypti* expresses a novel member of the α-amylase gene family. Insect Mol. Biol.**1:** 223–232.

Hagenbuchle, O., Bovery, R. & Young, R. A. 1980. Tissue-specific expression of mouse α-amylase genes: nucleotide sequence of isoenzyme mRNAs from pancreas and salivary gland. Cell **21:** 179–187.

Ibarrola, I. I., Navarro, E. & Urrutia, M. B. 2000. Acute and acclimated digestive responses of the cockle *Cerastoderma edule* (L.) to changes in food quality and quantity. I. Feeding and absorption of biochemical components. J. Exp. Mar. Biol. Ecol. **252:** 181–198.

Inomota, N., Kanda, K., Cariou, M. L. Tachida, H. & Yamazaki, T. 1995. Evolution of the response patterns to dietary carbohydrates and the developmental differentiation of gene expression of α-amylase in *Drosophila*. J. Mol. Evol. **41:** 1076–1084.

Janecek, S. 1994. Sequence similarities and evolutionary relationships of microbial, plant and animal α-amylases. Eur. J. Biochem. **224:** 519–524.

Janecek, S. 1997. α-Amylase family: molecular biology and evolution. Prog. Biophys. Mol. Biol. **67:** 67–97.

Le Moine, S., Sellos, D., Moal, J., Daniel, J. Y., San Juan Serrano, F., Samain, J. F. & Van Wormhoudt, A. 1997. α-Amylase in *Pecten maximus* (Mollusca, bivalves): protein and cDNA characterization; quantification of the expression in the digestive gland. Mol. Mar. Biol. Biotechnol. **6:** 228–237.

Le Moullac, G., Klein, B., Sellos, D. & Van Wormhoudt, A. 1996. Adaptation of trypsin, chymotrypsin and α-amylase to casein level and protein source in the shrimp *Penaeus vannamei*. J. Exp. Mar. Biol. Ecol. **208:** 107–125.

Moal, J., Samain, J. F., Daniel, Y., Boudry, P. Bougrier, S., Sellos, D. & Van Wormhoudt, A. 2000a. Evidence of absorption efficiency differences in two subpopulations of *Crassostrea gigas*. J. Shellfish Res. **19:** 616.

Moal, J., Daniel, J. Y., Sellos, D., Van Wormhoudt, A. & Samain, J. F. 2000b. Amylase mRNA expression in *Crassostrea gigas* during feeding cycles. J. Comp. Physiol. **170B:** 21–26.

Ohashi, K., Natori, S. & Kubo, T. 1999. Expression of amylase and glucose oxidase in the hypopharyngeal gland with an age-dependent role change of the worker honeybee (*Apis mellifera* L.). Eur. J. Biochem. **265:** 127–133.

Thompson, J. D., Higgins, D. G. &Gibson, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrice choice. Nucleic Acids Res. **22:** 4673–4680.

Unger, M. E. & Roesijadi, G. 1993. Sensitive assay for molluscan metallothinein induction based on ribonuclease protection and molecular titration of metallothionein and actin mRNA. Mol. Mar. Biol. Biotechnol. **2:** 319–324.

Van Wormhoudt, A. & Sellos, D. 1996. Cloning and sequencing analysis of three α-amylase cDNAs in the shrimp *Penaeus vannamei* (Crustacea Decapoda), evolutionary aspects. J. Mol. Evol. **42:** 543–551.

Received October 4, 2001
Accepted May 3, 2002