Identification of homologues of the mammalian intestinal lactase gene in non-mammals (birds and molluscs)

Jean-Noël FREUND*, Bernard JOST, Olivier LORENTZ and Isabelle DULUC

INSERM Unité 381, 3 avenue Molière, 67200 Strasbourg, France

Mammalian intestinal lactase hydrolyses a variety of β -glycosides and is processed from a precursor comprising four tandem domains exhibiting sequence similarity, suggestive of multiple duplication events in the evolutionary past. The aim of the present study was to investigate whether genes homologous to the lactase gene exist in animals other than mammals. A reverse transcriptase-PCR strategy using a degenerate mixture of oligonucleotides was developed to search for the presence of transcripts similar in sequence to the mammalian lactase mRNA in the digestive tracts of a bird (the chicken) and an invertebrate (the mussel). Partial cDNAs corresponding to the 3' end of intestinal mRNAs were identified in both animals. In chicken, two cDNAs were isolated, corresponding to 6.5 kb transcripts that used two distinct polyadenylation sites. In mussels, three cDNAs were

obtained and classified into two categories. One class of cDNA hybridized to a major mRNA of 3.5 kb and to minor species of 4.5 kb and 6 kb. The second class of cDNA hybridized to a 13 kb transcript, which was approximately twice as large as the mammalian lactase mRNA. Peptide sequences predicted from the chicken and mussel cDNAs confirmed that the proteins are related to mammalian lactase. They also suggested that the chicken protein and one mussel protein are integral molecules anchored in the cell membrane by a C-terminal transmembrane anchor, like lactase. These data provide evidence that proteins phylogenetically related to the mammalian-specific lactase are widespread in the animal kingdom, and that these proteins are expressed in the intestinal tract.

INTRODUCTION

Mammals are characterized by the ability of mothers to produce milk in the mammary gland, and by the ability of neonates to synthesize intestinal enzymes required for milk digestion. In particular, the high level of lactose contained in milk is hydrolysed into glucose and galactose by the mammalian-specific lactase, also termed lactase-phlorizin hydrolase (LPH; EC 3.2.1.62/108) or the neutral β -galactosidase complex. LPH, encoded by the *LCT* (lactase-encoding) gene, is a bifunctional enzyme endowed with two enzyme activities: the typical lactase activity and a phlorizin hydrolase activity involved in the digestion of glycosylceramides [1]. The 130–160 kDa mature form of LPH, anchored in the brush border membrane of the enterocytes, is synthesized as a 230–250 kDa single-chain precursor [2–5]. The precursor is composed of three parts: a signal peptide at the N-terminus, a transmembrane anchor followed by a cytoplasmic tail at the C-terminus, and a large central region subdivided into four repeated domains in tandem, defined by a fourfold inner sequence similarity [6–8]. Thus the molecular organization of the precursor suggests that the *LCT* gene has evolved by two cycles of internal duplication [6]. Mature LPH corresponds to the C-terminal half of the precursor, and includes the third and fourth domains, the hydrophobic anchor and the cytoplasmic tail.

Whereas lactase proper is specific to mammals, the ability to hydrolyse phlorizin {1-[2-(β -D-glycopyranosyloxy)-4,6-dihydroxyphenyl]-3-(4-hydroxyphenyl)-1-propanone} has been reported in many vertebrates [9,10]. Interestingly, we have shown that mammalian LPH also hydrolyses laminaribiose (3-*O*- β -D-glycopyranosyl-D-glucose), a constituent of the plant cell wall [11], like the laminaribiase present in the intestines of fish, insects and molluscs. These observations, as well as the duplicated

structure of the *LCT* gene, leave open the possibility of an ancient origin for this mammalian-specific gene. β -Glucosidase A of *Clostridium thermocellum* has been shown to exhibit sequence similarity with mammalian LPH [12]. β -Glucosidase A is the prototype of enzymes gathered together in the β -glucosidase A family, also termed the type 1 family of glycosylhydrolases (T1GH family). The T1GH family comprises single-domain β -glucosidases, phospho- β -glucosidases and β -galactosidases of eubacteria and archaeobacteria (see [12,13] and references therein), single-domain β -glucosidases belonging to defence systems or involved in cell wall β -glucan metabolism in plants [14–16], and the unique four-domain LPH precursor of mammals.

Genes related to the *LCT* gene have not been reported in animals other than mammals. In the present study we have developed a reverse transcriptase-PCR (RT-PCR) strategy using oligo(dT) and a degenerate oligonucleotide mixture to investigate whether homologues of the *LCT* gene are present in a non-mammalian vertebrate (chicken; *Gallus gallus*) and an invertebrate animal (mussel; *Mytilus edulis*).

MATERIAL AND METHODS

Oligonucleotides

Oligonucleotides were synthesized chemically (Eurogentec, Seraing, Belgium). The sequences of the mixture of 256 27-base oligonucleotides termed LPHOMOL are given in Figure 1. The other oligonucleotides used for this study were: ML1.1, 5'-TTATCAATACATCAATGCCTTATAT-3'; ML2.1, 5'-TTATGTATCAAAGCGAAGGGAAACA-3'; ML3.1, 5'-ATTCCTAAATACCCCTTTAATCTTT-3'.

Oligonucleotides (25 pmol) were labelled at the 5' end by kination for 1 h at 37 °C using [γ - 32 P]ATP (150 TBq/mmol;

Abbreviations used: *LCT* gene, lactase-encoding gene; LPH, lactase-phlorizin hydrolase; LPHOMOL, mixture of 256 27-base oligonucleotides (see Figure 1); RT-PCR, reverse transcriptase-PCR; T1GH, type 1 glycosylhydrolase.

* To whom correspondence should be addressed.

The nucleotide sequence data for the plasmids reported in this paper have been deposited in the EMBL/GenBank/DBJ Nucleotide Sequence Databases under accession numbers Z68488 (pGLH1), Z68489 (pGLH2), Z68490 (pMLH1), Z68491 (pMLH2) and Z68492 (pMLH3).

Figure 1 Synthetic oligonucleotide mixture LPHOMOL

(A) Amino acid sequence alignment of a region that is conserved between *Clostridium thermocellum* β-glucosidase A and the repeated domains of the human LPH precursor. The numbers refer to the positions of these regions in the respective protein sequences. (B) Nucleotide sequence of LPHOMOL, the mixture of 256 27-base oligonucleotides derived from the protein region conserved between *Clostridium thermocellum* β-glucosidase A and the human LPH precursor.

ICN, Orsay, France) and 5 units of T4 polynucleotide kinase (Boehringer, Mannheim, Germany) in 50 µl of 50 mM Tris/HCl, pH 7.6, 10 mM MgCl₂, 5 mM dithiothreitol and 0.1 mM spermidine. They were purified by electrophoresis on 20% (w/v) polyacrylamide gels, and recovered by elution in water.

RNA preparation and Northern blots

The mid-jejunum of a chicken and of an adult rat, as well as the whole intestines of 50 wild Breton mussels (*Mytilus edulis*), were dissected out and carefully rinsed with cold saline buffer. For chicken and rat, the intestinal mucosa was scraped with glass slides. Cytoplasmic RNA was extracted using the lithium chloride/urea method [18]. Polyadenylated RNA from mussel intestines was prepared by affinity chromatography on oligo(dT)-cellulose columns [17].

Samples containing 25 µg of cytoplasmic RNA or 2.5 µg of polyadenylated RNA were fractionated by electrophoresis on denaturing 1% agarose/17% formaldehyde gels, and transferred to nitrocellulose filters [17]. The filters were prehybridized for 3 h at 42 °C in 50% deionized formamide, 5× SSC (1× SSC is 150 mM NaCl and 15 mM sodium citrate), 0.2% Ficoll, 0.2% polyvinylpyrrolidone, 100 mM phosphate buffer, pH 6.5, 0.1% SDS and 100 µg/ml denatured calf thymus DNA. They were then incubated overnight at 42 °C in the prehybridization solution supplemented with 10% (w/v) dextran sulphate and 0.2 pmol/ml radiolabelled PCR-amplified DNA probe (see below). The filters were washed for 2×30 min in 2× SSC/0.1% SDS at 25 °C, 2×15 min in 0.2× SSC/0.1% SDS at 55 °C and 2×15 min in 0.2× SSC/0.1% SDS at 60 °C, and were then exposed to autoradiography.

DNA preparation and Southern blots

Genomic DNA was extracted from the coat of a single wild Breton mussel (Genomix, Euromedex, France). Plasmid DNA was prepared by the alkali method and purified by centrifugation on caesium chloride gradients [17].

DNA was fractionated by electrophoresis on 1% or 3% agarose gels, denatured in 0.5 M NaOH/1.5 M NaCl and transferred to nitrocellulose filters (Schleicher & Schuell, Kassel, Germany) using standard procedures [17]. The filters were incubated for prehybridization for 3 h at 42 °C in 10% deionized formamide, 5× SSC, 0.2% Ficoll, 0.2% polyvinylpyrrolidone, 35 mM phosphate buffer, pH 6.5, 0.1% SDS and 100 µg/ml denatured calf thymus DNA. They were then placed overnight at

42 °C in prehybridization solution to which 1 pmol/ml radiolabelled oligonucleotide had been added. The filters were washed for 2×30 min in 2× SSC/0.1% SDS at 25 °C, 2×15 min in 0.2× SSC/0.1% SDS at 42 °C and 2×15 min in 0.1× SSC/0.1% SDS at 42 °C, and then exposed to autoradiography.

RT-PCR strategy

Single-stranded cDNAs were synthesized at 42 °C for 1 h from 5 µg of cytoplasmic RNA in 25 µl of solution containing 50 pmol of 20-mer oligo(dT), 50 mM Tris/HCl, pH 8.3, 50 mM KCl, 10 mM MgCl₂, 10 mM dithiothreitol, 0.5 mM spermidine, 4 mM sodium pyrophosphate, 0.2 mM of each dNTP and 15 units of avian myeloblastosis reverse transcriptase (Promega, Madison, WI, U.S.A.).

The PCR was carried out with 3 µl of the solution of single-stranded cDNAs, using 75 pmol of LPHOMOL and 75 pmol of oligo(dT). The reaction was performed in 100 µl of 75 mM Tris/HCl, pH 9, 20 mM (NH₄)₂SO₄, 0.01% Tween 20, 1.25 mM MgCl₂, 0.2 mM of each dNTP and 0.5 unit of DNA polymerase from a *Thermus* species (GoldStar polymerase; Eurogentec). DynaWax (Eurogentec) was used to separate the cDNA template and the primers from the polymerase during pipetting and before the first step of denaturation. The cycling parameters were as follows: denaturation at 94 °C for 30 s, hybridization at 45 °C for 30 s and polymerization at 72 °C for 30 s. Amplification was conducted for 45 cycles in a MicroCycler E (Eppendorf, Hamburg, Germany).

In control experiments, PCR amplification of oligo(dT)-primed single-stranded cDNAs was performed using only one oligonucleotide, i.e. LPHOMOL or oligo(dT). In other cases, the cDNA synthesis was omitted and the PCR was carried out directly with 1 µg of RNA. Alternatively, the PCR was conducted in the absence of any source of cDNA or RNA.

For radiolabelling, 2 ng of cDNA was processed for PCR during eight cycles in 50 µl of solution containing 75 mM Tris/HCl, pH 9, 20 mM (NH₄)₂SO₄, 0.01% Tween 20, 1.25 mM MgCl₂, 0.05 mM each of dATP, dGTP and dTTP, 1 MBq of dCTP (110 TBq/mmol; ICN Biomedicals), 2 pmol of LPHOMOL, 2 pmol of oligo(dT) and 0.5 unit of GoldStar DNA polymerase (Eurogentec). Labelled DNA was separated from the unincorporated nucleotides by chromatography on a Sephacryl S200 HR column (Pharmacia).

Cloning and sequencing of RT-PCR products

RT-PCR products were loaded on a Centricon-100 column (Amicon, Grace Company, Epervan, France) and centrifuged at 1000 g for 20 min. The cDNA retained in the upper reservoir was precipitated with ethanol. Recombinant plasmids were obtained by ligation of 300 ng of cDNA and 50 ng of pGEM-T vector (Promega). The nucleotide sequences of the cDNA fragments were determined on both strands by the dideoxy method (T7 sequencing kit; Pharmacia). Nucleotide and protein sequences were analysed with PCGENE software (Intelligenetics, Oxford, U.K.).

RESULTS

Evaluation of the RT-PCR strategy with oligo(dT) and LPHOMOL using rat intestinal RNA

To investigate whether genes related to the *LCT* gene (or belonging to the T1GH family) exist in animals other than mammals, we have developed an RT-PCR strategy using as amplification primers oligo(dT) and a mixture of 256 27-base oligonucleotides, termed LPHOMOL. The sequence of

Figure 2 RT-PCR of rat intestinal RNA using LPHOMOL and oligo(dT)

PCR was performed on oligo(dT)-primed cDNA derived from rat intestinal RNA, using LPHOMOL and oligo(dT) as amplification primers (lane 1). The expected cDNA fragment of 780 bp appeared within a smear of DNA. Controls were the following: PCR was carried out on intestinal RNA without the reverse transcription step (lane 2), or it was performed on oligo(dT)-primed cDNA using LPHOMOL (lane 3) or oligo(dT) (lane 4) as a single amplification primer. In another experiment, PCR was performed with LPHOMOL and oligo(dT) in the absence of any source of RNA or cDNA template (lane 5). In the controls, except lane 4, a band of 129 bp was obvious.

Figure 3 Identification of LPH-related transcripts in chicken intestine by RT-PCR and Northern blotting

(A) PCR was carried out on oligo(dT)-primed cDNA derived from chicken intestinal RNA (lane 1) or liver RNA (lane 3), using LPHOMOL and oligo(dT) as amplification primers. Two amplified cDNA fragments of 0.65 kb and 1.15 kb appeared for RNA extracted from the intestine. In the liver (lane 3) and in controls in which PCR was performed directly on intestinal RNA without the reverse transcription step (lanes 2 and 4 respectively), the experiment resulted in the amplification of a band of 129 bp. (B) Cytoplasmic RNA (30 μ g) extracted from chicken intestine was hybridized to radiolabelled cDNA probes synthesized from plasmids pGLH1 and pGLH2 (lane 1). In lane 2, rat intestinal RNA was hybridized to rat LPH cDNA. The positions of 28 S and 18 S rRNAs are indicated.

LPHOMOL was chosen according to three criteria (Figure 1): (1) it was derived from a peptide region conserved between *Clostridium thermocellum* β -glucosidase A and domains III and IV of mammalian LPH; (2) this region is located near the C-terminal end of the TIGH domains in the prokaryotic and mammalian proteins; and (3) the three nucleotides at the 3' end of LPHOMOL corresponded to a conserved amino acid represented by a single codon, TGG, encoding tryptophan.

RT-PCR with LPHOMOL and oligo(dT) was conducted using rat intestinal RNA to validate the method and to establish optimal PCR conditions. As expected, a 780 bp RT-PCR product was obtained with rat RNA samples (Figure 2, lane 1). Cloning of this cDNA fragment and sequencing of both extremities indicated that it actually resulted from amplification of the 3'-end of the LPH mRNA. This fragment was always included within a smear of DNA, and it was at least 20 times less abundant than the RT-PCR product obtained for the same

mRNA using a pair of non-degenerate oligonucleotides (results not shown). This was most probably a consequence of the relatively low sequence specificity of both upstream (LPHOMOL) and downstream [oligo(dT)] primers used in this study. In controls in which the cDNA synthesis step was omitted, a 129 bp DNA fragment was visualized (Figure 2, lane 2). This fragment was also produced when cycling was conducted with LPHOMOL as a unique primer, and in the absence of any source of RNA or cDNA template (Figure 2, lanes 3 and 5). However, it was never observed when a specific cDNA target was added in the PCR mixture. This 129 bp fragment contained a single copy of LPHOMOL at each extremity, flanking a central region that was not related to either LPHOMOL or the *LCT* gene sequence (results not shown).

Identification of LPH-related transcripts in the intestine of a bird, the chicken

To search for the presence of RNA transcripts related in sequence to the *LCT* gene in non-mammalian vertebrates, the above RT-PCR strategy was applied to samples extracted from the intestine of a bird, the chicken. It is well known that, in chicken, like in other birds, typical lactase activity is virtually absent [19,20]. The chicken was chosen because it has been the subject of many investigations in the field of intestinal ontogeny.

RT-PCR performed with oligo(dT) and LPHOMOL on chicken intestinal RNA resulted in the amplification of two major DNA fragments of 0.65 kb and 1.15 kb (Figure 3A, lane 1). However, no specific DNA product was visualized using the same conditions with RNA extracted from the liver (Figure 3A, lane 3). In the liver, and in intestinal controls in which the cDNA synthesis step was omitted, the 129 bp fragment previously described for control rat RNA was obvious (Figure 3A, lanes 2 and 4; see also Figure 2).

The DNA fragments of 0.65 kb and 1.15 kb were inserted into the pGEM-T vector to give the plasmids pGLH1 and pGLH2 respectively. Figure 4 shows the nucleotide sequence of these fragments and the deduced amino acid sequence. As expected, both inserts exhibited at one extremity the sequence of LPHOMOL, and at the opposite end a poly(A) tail preceded by a canonical polyadenylation signal, AATAAA. The strict identity between the first 627 bp of pGLH2 and the whole nucleotide sequence of pGLH1 [except for the two nucleotides immediately upstream of the poly(A) track] indicates that both cDNAs were derived from the same gene, and that they corresponded to two transcripts differing by the alternative use of two distinct polyadenylation sites.

Downstream of LPHOMOL, the cDNA fragments of pGLH1 and pGLH2 exhibited a 375 bp open reading frame. A comparison of the amino acid sequence predicted from this open reading frame with the C-terminal ends of *Clostridium thermocellum* β -glucosidase A and of domains III and IV of mammalian LPH indicated some degree of similarity. In particular, typical amino acids for a new member of the TIGH family were present at the expected positions in the chicken peptide sequence (Figure 4; see also Figure 7). This demonstrated that the intestinal chicken cDNAs isolated in the present study correspond to a new gene that is phylogenetically related to the mammalian *LCT* gene. It should also be emphasized that, downstream of the C-terminal end of the typical TIGH domain, the chicken protein exhibited a fragment of 85 residues that showed clear sequence similarity with the 87-amino-acid segment of the C-terminal part of mammalian LPH (see Figure 7). In particular, this fragment contained a stretch of 19 hydrophobic residues, located at the same position as the transmembrane anchor of mammalian LPH

Figure 4 Nucleotide and deduced amino acid sequences of chicken cDNAs inserted in plasmids pGLH1 and pGLH2

Nucleotides in lower-case characters correspond to LPHOMOL. Dashes indicate identical nucleotides in pGLH1 and pGLH2. The polyadenylation signals are underlined. Typical amino acids conserved in the T1GH family are boxed. The hydrophobic segment corresponding to the transmembrane anchor is double-underlined, and the conserved serine residue in the cytoplasmic tail is indicated by an arrowhead. The translation stop codon is indicated by ■. The accession numbers of the cDNA sequences of pGLH1 and pGLH2 in the EMBL database are Z68488 and Z68489 respectively.

Figure 5 Identification of LPH-related transcripts in mussel intestine by RT-PCR

(A) PCR conducted on oligo(dT)-primed cDNA from mussel intestine resulted in the amplification of three bands of 580, 350 and 270 bp when LPHOMOL and oligo(dT) were used as primers (lane 2). Only the band of 129 bp was obtained when the reverse transcription step was omitted (lane 1). In lanes 3, 4 and 5, the cDNA fragments inserted in plasmids pMLH1, pMLH2 and pMLH3 respectively were amplified by PCR using LPHOMOL and oligo(dT) as primers. In (B), (C) and (D), the same samples as in (A) were transferred to nitrocellulose filters and hybridized to the radiolabelled oligonucleotides ML1.1, ML2.1 and ML3.1 respectively.

[6,8]. This stretch of 19 amino acids was identified as a typical transmembrane spanning segment in the chicken protein. Downstream of this hydrophobic segment, the cytoplasmic tail contained a conserved serine, analogous to a possible target for *in vivo* phosphorylation in the mammalian LPH sequence [21].

To identify the transcripts corresponding to the cDNAs inserted in pGLH1 and pGLH2, cytoplasmic RNA extracted from the chicken intestine was separated by gel electrophoresis and hybridized to radiolabelled PCR probes synthesized from both plasmids. As shown in Figure 3(B), the nucleotide probes revealed transcripts of 6.5 kb that were nearly identical in size to rat LPH mRNA.

Identification by RT-PCR of transcripts homologous to mammalian LPH mRNA in mussel intestine

The finding of a gene homologous to the mammalian *LCT* gene in a non-mammalian vertebrate, i.e. the chicken, and the fact

that, as in mammals, this gene is transcribed in the intestinal tract, led us to hypothesize (i) that gene members of the T1GH family may be widespread in the animal kingdom, and (ii) that the intestine could represent a typical site for their expression. To address this question, we have applied to an invertebrate the same RT-PCR strategy as that developed in rat and chicken. For this purpose, cytoplasmic RNA was extracted from the intestine of a mollusc, the mussel (*Mytilus edulis*). Molluscs were chosen because they express laminaribiase in their intestine and because we have previously shown that laminaribiase activity present in the intestine of mammals is related to LPH [11].

LPHOMOL and oligo(dT) were used for RT-PCR on mussel intestinal RNA under the conditions established for the rat samples. Three fragments of 580 bp, 350 bp and 270 bp were detected within a high background of DNA (Figure 5A, lane 2). Except for the 129 bp fragment previously described in controls, no amplification product was found when the cDNA synthesis step was omitted (Figure 5A, lane 1). In addition, no specific RT-

Figure 6 Nucleotide and deduced amino acid sequences of mussel cDNAs inserted in plasmids pMLH1, pMLH2 and pMLH3

Nucleotides in lower-case characters correspond to LPHOMOL. Polyadenylation signals are underlined. Arrows underline sequences complementary to the synthetic oligonucleotides MLH1.1, MLH2.1 and MLH3.1. Typical amino acids conserved in the T1GH family are boxed. The hydrophobic segment representing the transmembrane anchor of the protein is double-underlined. The translation stop codon is indicated by ■. The accession numbers of the sequences of pMLH1, pMLH2 and pMLH3 in the EMBL database are Z68490 and Z68491 and Z68492 respectively.

Figure 7 Protein sequence alignment of *Clostridium thermocellum* β-glucosidase A (β-Gluc A), domains of the human LPH precursor (LPH dII, dIII, dIV), and the chicken (GLH1/2) and mussel proteins (MLH1–MLH3) corresponding to cDNAs identified in this study

Typical amino acids of members of the T1GH family are boxed (according to [12]). In the T1GH domain, identical amino acids or amino acids belonging to the same chemical family and present in at least three out of the eight sequences are in bold. Downstream of the T1GH domain, amino acids common to two sequences are in bold. The hydrophobic transmembrane segments are underlined. Gaps have been introduced in the sequences for maximum alignment (–). Translation stop codons are indicated by ■.

PCR product was obtained with RNA extracted from another organ, i.e. the coat of the mussels (results not shown). The RT-PCR products obtained from mussel intestinal RNA were inserted into the pGEM-T vector, and three distinct plasmids were isolated. These plasmids, named pMLH1, pMLH2 and pMLH3, contained DNA inserts of 212 bp, 395 bp and 518 bp respectively, followed by a poly(A) tail of 40–60 residues. Figure 6 shows the nucleotide sequences of the mussel cDNA fragments, and the deduced amino acid sequences. All the inserts exhibited a canonical polyadenylation signal (AATAAA or AATAGA) 15–16 nucleotides upstream of the poly(A) tail. Sequences at the

other extremity began with LPHOMOL and continued with an open reading frame extending over 156 bp for pMLH1 and pMLH2, and over 288 bp for pMLH3. The 3' untranslated sequences were completely different for the three cDNAs, providing the opportunity to synthesize specific oligonucleotide probes for each recombinant plasmid: ML1.1 for pMLH1, ML2.1 for pMLH2 and ML3.1 for pMLH3 (see Figure 6). These oligonucleotides were used to characterize definitively the cDNA inserts of pMLH1, pMLH2 and pMLH3 with respect to the three RT-PCR products obtained from mussel intestinal RNA. For this purpose, RT-PCR products obtained using LPHOMOL

Figure 8 Mussel genomic DNA analysis

PCR was conducted with plasmid pMLH1 (lane 1), pMLH2 (lane 2) or pMLH3 (lane 3), or with mussel genomic DNA (lanes 4–6) using the oligonucleotide pairs ML1.1/LPHOMOL (lanes 1 and 4), ML2.1/LPHOMOL (lanes 2 and 5) and ML3.1/LPHOMOL (lanes 3 and 6). With these oligonucleotide pairs, PCR resulted in the amplification of DNA fragments of 192 bp, 189 bp and 497 bp using either the plasmids or the genomic DNA as template.

and oligo(dT), and PCR fragments obtained using the same oligonucleotide pair and the template plasmids pMLH1–pMLH3, were separated by gel electrophoresis and hybridized to the radiolabelled oligonucleotides ML1.1, ML2.1 and ML3.1 (Figures 5B–5D). ML1.1 revealed the 270 bp RT-PCR product, and it selectively hybridized to the cDNA fragment resulting from plasmid pMLH1. No cross-hybridization was observed with the PCR fragments obtained with pMLH2 and pMLH3. Similarly, ML2.1 and ML3.1 specifically hybridized to the RT-PCR fragments of 350 bp and 580 bp respectively, and to the PCR products from plasmids pMLH2 and pMLH3 respectively. Thus we concluded that the cDNA inserts of pMLH1, pMLH2 and pMLH3 respectively corresponded to the 270 bp, 350 bp and 580 bp fragments obtained by RT-PCR on mussel intestinal RNA.

As reported above in the case of the chicken, the amino acid sequences predicted from the open reading frames of pMLH1, pMLH2 and pMLH3 exhibited similarity with the C-terminal ends of *Clostridium thermocellum* β -glucosidase A and of LPH domains III and IV, in particular at the level of typical residues that characterize members of the T1GH family (Figures 6 and 7). However, despite the similarity at the amino acid level, the nucleotide sequences of pMLH1, pMLH2 and pMLH3 showed clear differences. Indeed, in addition to the 3' untranslated region already mentioned, point mutations were observed between the open reading frames of pMLH1 and pMLH2, resulting in a single amino acid substitution. In addition, the nucleotide sequence of pMLH3 was unambiguously different from those of pMLH1 and pMLH2 at the level of the 3' untranslated region, as well as at the level of the open reading frame. The amino acid sequence predicted from pMLH3 showed a 44-amino-acid extension at the C-terminal end as compared with the sequences deduced from pMLH1 and pMLH2. This stretch of amino acids comprises a hydrophobic segment of 20 residues immediately upstream of the C-terminal end that might constitute a transmembrane helix anchor for a protein classified as an integral protein. However, it should be emphasized that, despite the functional homology, the hydrophobic transmembrane spanning segments of mammalian LPH and of the mussel protein do not exhibit any primary sequence similarity (Figure 7). The fact that one class of mussel protein contained a C-terminal anchor (the class corresponding to pMLH3), whereas another class did not (that corresponding to

Figure 9 Northern blot analysis of mussel intestinal RNA

A sample of 2.5 μ g of polyadenylated RNA was loaded in each lane and hybridized to cDNA probes prepared from plasmid pMLH1 (lane 1), pMLH2 (lane 2) or pMLH3 (lane 3).

pMLH1 and pMLH2), suggests a different role and/or a different tissue distribution of these proteins in the mussel intestine.

Evidence for the presence of genes homologous to the mammalian *LCT* gene in the genome of the mussel

Given (1) the intrinsic efficiency of the RT-PCR technology and (2) the fact that the present study was conducted with a pair of oligonucleotides exhibiting a relatively low degree of sequence specificity, i.e. oligo(dT) and the LPHOMOL mixture of 256 oligonucleotides, we asked whether the three cDNAs isolated from 'mussel RNA' actually corresponded to mussel genes. The rationale of this question was related to the fact that intestinal epithelial cells *in vivo* may be in close vicinity to symbiotic micro-organisms, which could themselves express protein members of the T1GH family.

To address this question, genomic DNA was extracted from the coat of the mussel, an organ devoid of symbiotic micro-organisms. DNA extraction was performed on a single animal, and the resulting genomic DNA was tested by PCR using LPHOMOL and the oligonucleotides ML1.1, ML2.1 and ML3.1. In parallel, PCR was carried out with plasmids pMLH1, pMLH2 and pMLH3, using the same primers. As shown in Figure 8, PCR products were obtained with mussel genomic DNA and every oligonucleotide pair. The fragments exhibited the same size as those produced with the template plasmids pMLH1, pMLH2 and pMLH3. These data demonstrate that the three cDNAs identified by RT-PCR in intestinal RNA actually correspond to genes present in the mussel genome. Since DNA extraction was performed on a single animal, they also suggest that mussels contain at least two, and possibly three, distinct genes related to the mammalian *LCT* gene.

Northern blot analysis of mussel LPH-related transcripts

To identify the transcripts corresponding to the three mussel cDNAs isolated in the present study, polyadenylated RNA extracted from mussel intestine was fractionated by gel electrophoresis and transferred to nitrocellulose filters. The filters were incubated with radiolabelled PCR probes synthesized from the plasmids pMLH1, pMLH2 and pMLH3, using LPHOMOL and oligo(dT) as primers. As shown in Figure 9, the probes derived from pMLH1 and pMLH2 hybridized to a major transcript of 3.5 kb and to two minor RNA species of 4.5 kb and 6 kb. The probe corresponding to pMLH3 did not cross-hybridize with any of the transcripts detected by pMLH1 or pMLH2. However, it hybridized to an mRNA whose size was estimated by gel electrophoresis to be around 13 kb.

DISCUSSION

By RT-PCR using oligo(dT) and the LPHOMOL mixture of 256 oligonucleotides derived from a region conserved between mammalian LPH and *Clostridium thermocellum* β -glucosidase A, we have identified a chicken gene that is phylogenetically related to the mammalian *LCT* gene, and two or three additional related genes in an invertebrate, the mussel. Like the *LCT* gene in mammals, the chicken and mussel lactase-related genes are expressed in the intestinal tract, although lactase activity is virtually absent from the intestines of birds and molluscs. These genes represent new members of the T1GH family, since typical amino acid residues that are conserved between protein members of this family are present in the amino acid sequences predicted from the chicken and mussel cDNAs. The presence of typical amino acids is confirmed by additional sequence data over 1.5 kb from cDNAs corresponding to the transcripts of 3.5 kb and 13 kb in mussels (results not shown). These additional sequences indicate (1) that the mussel proteins are organized into tandem repeats of the prototype T1GH domain, and (2) that the mussel proteins contain the typical Glu residues identified as being present in the active sites of mammalian LPH and other members of this family [8,22–24].

Evidence is provided for the first time in the present study that genes phylogenetically related to the mammalian *LCT* gene are widespread in the animal kingdom. From the physiological point of view, the fact that these genes are expressed in adult birds and molluscs may provide an explanation as to why the *LCT* gene is transcribed in adult mammals, although they no longer consume lactose after weaning and lactase activity is down-regulated at this stage [25]. Indeed, the maintenance of active expression throughout life in mammals may be reminiscent of a possible nutritional requirement for an enzyme of the T1GH family at the adult stage in mammalian ancestors, independent of the presence of lactose in the diet. The substrates of the chicken and mussel enzymes are not known at present. However, we have observed specific staining of the mussel intestinal epithelium, and of intraluminal structures, using *X*- β -galactoside or *X*- β -fucoside as chromogenic substrates (J.-N. Freund, B. Jost, O. Lorentz and I. Duluc, unpublished work). The characterization of the natural substrates, together with determination of the primary structures of the enzymes, would provide further insight into the mechanisms involved in substrate specificity and catalysis. Our present results should also be borne in mind with regard to the absence of lactase activity from the intestinal brush border membrane of specific mammals, i.e. pinipeds and macropods [26,27]. Although this absence could reflect a lack of expression of any protein of the T1GH family in these mammals, the present results suggest the possibility that changes in substrate specificity as a result of point mutations may have prevented the enzymic detection of a protein that is still present.

The results reported here open the way to further investigations in two major directions. The first concerns the evolution of members of the T1GH family in animals, with regard to the number of domains, processing and structure–function relationship of these proteins. In this field, it should be emphasized that all the members of the family described so far are single-domain enzymes, with the exception of the four-domain LPH precursor. Interestingly, one can note that the 3.5 kb transcript identified in mussels is long enough to code for a two-domain protein, while the 13 kb mRNA may encode a protein with eight domains. Another point concerns the identity of the active sites within the multimeric enzymes, since conflicting results have been reported for the location of the lactase and phlorizin hydrolase sites in LPH [8,24]. Also intriguing is the role of the ‘pro’ part of the

LPH precursor, as some data suggest that it may be involved in the folding and/or sorting of the mature enzyme ([28,29]; B. Jost, I. Duluc and J. N. Freund, unpublished work). The second direction concerns the regulation of members of the T1GH family in the animal kingdom, since these enzymes show intestinal expression in invertebrates, birds and mammals. It should be emphasized that the transcription of the mammalian *LCT* gene is under the control of a homoeoprotein belonging to the *caudal* gene family [30], and that genes of this family are known to be expressed in the intestines of invertebrates, birds and mammals [31–34]. Hence members of the T1GH family in animal species that are convenient for thorough genetic and molecular studies may represent attractive candidates as targets of nuclear factors of the Caudal type, in order to investigate the role of homoeobox genes in the process of intestinal ontogeny and cellular differentiation.

Note added in proof (received 17 January 1997)

A single-domain cytoplasmic member of the T1GH family has recently been reported in the liver of mammals [35].

REFERENCES

- 1 Semenza, G. and Auricchio, S. (1989) in *The Metabolic Basis of Inheritable Disease*, (Scriver, C. R., Beaudet, A. L., Sly W. S. and Valle, D., eds.), pp. 2975–2997, McGraw Hill, New York
- 2 Skovbjerg, H., Danielsen, E. M., Noren, O. and Sjöström, H. (1984) *Biochim. Biophys. Acta* **798**, 247–251
- 3 Büller, H. A., Montgomery, R. K., Sasak, V. W. and Grand, R. J. (1987) *J. Biol. Chem.* **262**, 17206–17211
- 4 Naim, Y. H., Sterchi, E. E. and Lentze, M. G. (1987) *Biochem. J.* **241**, 427–434
- 5 Naim, H., Lacey, S., Sambrook, J. F. and Gething, M. J. H. (1991) *J. Biol. Chem.* **266**, 12313–12320
- 6 Mantei, N., Villa, M., Enzler, T., Wacker, H., Boll, W., James, P., Hunziker, W. and Semenza, G. (1988) *EMBO J.* **7**, 2705–2713
- 7 Duluc, I., Boukamel, R., Mantei, N., Semenza, G., Raul, F. and Freund, J.-N. (1991) *Gene* **103**, 275–276
- 8 Wacker, H., Keller, P., Falchetto, R., Legler, G. and Semenza, G. (1992) *J. Biol. Chem.* **267**, 18744–18752
- 9 Colombo, V., Lorentz-Meyer, H. and Semenza, G. (1973) *Biochim. Biophys. Acta* **327**, 412–424
- 10 Leese, J. H. and Semenza, G. (1973) *J. Biol. Chem.* **248**, 8170–8173
- 11 Freund, J.-N., Gossé, F. and Raul, F. (1991) *Enzyme* **45**, 71–74
- 12 Gräbnitz, F., Seiss, M., Rücknagel, K. P. and Staudenbauer, W. L. (1991) *Eur. J. Biochem.* **200**, 301–309
- 13 Henrissat, B. A. (1991) *Biochem. J.* **280**, 309–316
- 14 Lenman, M., Falk, A., Xue, J. and Rask, L. (1993) *Plant Mol. Biol.* **21**, 463–474
- 15 Gus-Mayer, S., Brunner, H., Schneider-Poetsch, H. A. W. and Rüdiger, W. (1994) *Plant Mol. Biol.* **26**, 909–921
- 16 Leah, R., Kigel, J., Svendsen, I. and Mundy, J. (1995) *J. Biol. Chem.* **270**, 15789–15797
- 17 Sambrook, J. F., Fritsch, E. E. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor
- 18 Auffray, C. and Rougeon, F. (1980) *Eur. J. Biochem.* **107**, 303–314
- 19 Kerry, K. R. (1969) *Comp. Biochem. Physiol.* **29**, 1015–1022
- 20 Zoppi, G. and Shmerling, D. H. (1969) *Comp. Biochem. Physiol.* **29**, 289–294
- 21 Keller, P., Semenza, G. and Shaltiel, S. (1995) *FEBS Lett.* **368**, 563–567
- 22 Tull, D., Withers, S., Gilkes, N. R., Kilburn, D. G., Warren, R. A. J. and Aebersold, R. (1991) *J. Biol. Chem.* **266**, 15621–15625
- 23 Trimbis, D. E., Warren, R. A. J. and Withers, S. (1992) *J. Biol. Chem.* **267**, 10248–10251
- 24 Neele, A. M., Einerhand, A. W. C., Dekker, J., Büller, H. A., Freund, J.-N., Verhave, M., Grand, R. J. and Montgomery, R. K. (1995) *Gastroenterology* **109**, 1234–1240
- 25 Duluc, I., Jost, B. and Freund, J.-N. (1993) *J. Cell Biol.* **123**, 1577–1586
- 26 Sunshine, P. and Kretschmer, N. (1964) *Science* **144**, 850–851
- 27 Crisp, E., Czolij, R. and Messer, M. (1987) *Comp. Biochem. Physiol.* **88B**, 923–927

- 28 Oberholtzer, T., Mantei, N. and Semenza, G. (1993) *FEBS Lett.* **333**, 127–131
- 29 Naim, Y. H., Jacob, R., Naim, H., Sambrook, J. F. and Gething, M. J. H. (1994) *J. Biol. Chem.* **269**, 26933–26943
- 30 Troelsen, J. T., Olsen, J., Micheltore, C., Hansen, G. H. and Noren, H. (1994) *FEBS Lett.* **342**, 297–301
- 31 Mlodzik, M. and Gehring, W. J. (1987) *Cell* **48**, 465–478
- 32 Duprey, P., Chowdhury, K., Dressler, G. R., Balling, R., Simon, D., Guenet, J.-L. and Gruss, P. (1988) *Genes Dev.* **2**, 1647–1654
- 33 Freund, J.-N., Boukamel, R. and Benazzouz, A. (1992) *FEBS Lett.* **314**, 163–166
- 34 Doll, U. and Niessing, J. (1993) *Dev. Biol.* **156**, 155–163
- 35 Hays, W. S., Jenison, S. A., Yamada, T., Pastuszyn, A. and Glew, R. H. (1996) *Biochem. J.* **319**, 829–837

Received 9 August 1996/10 October 1996; accepted 28 October 1996