Molecular evolution of protein O-fucosyltransferase genes and splice variants

C. Loriol2, F. Dupuy2, R. Rampal3, M.A. Dlugosz3, R.S. Haltiwanger3, A. Maftah1,2, and A. Germot2

1INRA, UMR 1061, Unité de Génétique Moléculaire Animale, GDR-CNRS 2590, Université de Limoges, Institut des Sciences de la Vie et de la Santé, Faculté des Sciences et Techniques, 87060 Limoges, France; and 2Department of Biochemistry and Cell Biology, Institute for Cell and Developmental Biology, State University of New York, Stony Brook, NY 11794

Received on February 3, 2006; revised on April 12, 2006; accepted on April 26, 2006

O-Fucose has been described on both epidermal growth factor-like (EGF-like) repeats and Thrombospondin type 1 repeats (TSRs). The enzyme adding fucose to EGF-like repeats, protein O-fucosyltransferase 1 (Pofut1), is a soluble protein located in the lumen of endoplasmic reticulum (ER). A second protein O-fucosyltransferase, Pofut2, quite divergent from its homolog Pofut1, has recently been shown to O-fucosylate TSRs but not EGF-like repeats. To date, Pofut1 genes have only been characterized in human, mouse, and fly, and Pofut2 in mouse, fly, and partially in the nematode Caenorhabditis elegans. Here, we report cDNA sequences and genomic structures of bovine Pofut1 and Pofut2 genes and describe for the first time five alternative spliced transcripts for each gene. Only one transcript for both Pofut1 and Pofut2 encodes an active bovine O-fucosyltransferase. Variant transcript distribution was examined in 13 bovine tissues. Transcripts encoding active forms are ubiquitous, whereas other forms possess a more restricted tissue-expression profile. Sequence comparison and phylogenetic analyses revealed that both Pofut genes are present as a single copy in animal genomes, and their exon–intron organizations are conserved among vertebrates. The last common ancestor of all analyzed bilateral species would be predicted to possess polyxenic Pofut genes in their genome.

Key words: alternative splicing/fucose/gene organization/O-fucosyltransferase activity/Pofut1 and Pofut2 phylogeny

Introduction

Fucose is commonly found as a terminal sugar residue at the nonreducing end of glycans in glycoproteins and glycosphingolipids of eukaryotic cells (Lochnit et al., 2001; Wittmann, 2001). It is also found directly O-linked to serine or threonine residues of proteins (Hallgren et al., 1975). Among these, urokinase-type plasminogen activator was the first described O-fucosylated protein (Kentzer et al., 1990). Subsequently, several other O-fucosylated proteins were identified including tissue-type plasminogen activator (Harris et al., 1991), factors VII, IX, and XII (Bojern et al., 1991; Harris et al., 1992; Nishimura et al., 1992), Notch (Moloney, Panin, et al., 2000, Moloney, Shair, et al. 2000), Delta/Serrate (Panin et al., 2002), and Cripto (Schiffer et al., 2001; Yan et al., 2002). Interestingly, the O-fucosylated residue in all of these proteins was found between the second and the third conserved cysteines within an epidermal growth factor EGF-like (EGF-like) repeat (Figure 1A). The consensus sequence, C²X₄₋₅(S/T)C³, for O-fucosylation was deduced from the comparison of glycosylation sites on these proteins (Harris and Spellman, 1993; Shao and Haltiwanger, 2003). The GDP-fucose protein O-fucosyltransferase 1 (Pofut1) adds O-fucose to EGF-like repeats and was originally identified in and purified from Chinese hamster ovary (CHO) cells (Wang et al., 1996; Wang and Spellman, 1998). Recent studies have revealed that O-fucosylation of EGF-like repeats occurs in the endoplasmic reticulum (ER) and that Drosophila Ofut1 has a distinct Notch chaperone activity (Luo and Haltiwanger, 2005; Okajima et al., 2005). The corresponding Pofut1 gene (Glyco-syltransferase family 65 in the CAZY [carbohydrate-active enzyme] database [Coutinho et al., 2003]) is widely present in metazoans (Wang and Spellman, 1998; Wang et al., 2001). O-Fucose on EGF-like repeats can be further elongated by members of the Fringe family, fucose-specific β1,3-N-acetylglucosaminyltransferases (Bruckner et al., 2000; Moloney, Panin, et al., 2000). The GlcNAc-β1,3-Fuc disaccharide can be further elongated by β1,4-galactosyltransferase 1 (Chen et al., 2001) and either α2,3- or α2,6-sialyltransferase (Moloney, Panin, et al., 2000) to form the tetrasaccharide: NeuAcα2,3-Galβ1,4-GlcNAc-β1,3-Fuc-O-S/T (Figure 1A). O-Fucosylglycans modulate the function of proteins by controlling their interactions. O-Fucosylation of EGF-like repeats is essential for the interaction of Notch receptor with its ligands, Delta and Serrate/Jagged (Okajima and Irvine, 2002; Sasamura et al., 2003; Shi and Stanley, 2003). Indeed, the down-regulation of Ofut1 gene expression by RNA interference (RNAi) in Drosophila, or by Pofut1 exon 2 deletion in mice, impairs Notch signaling. Mouse embryos lacking Pofut1 die at day 9.5 postcoitum with severe embryonic defects (Shi and Stanley, 2003). The phenotypes of Drosophila and mouse knockout mutants for Notch receptors resemble those obtained by the reduction or abolition of Pofut1 gene expression (Barrantes et al., 1999; Krebs et al., 2000; Lutolf et al., 2002; Okajima and Irvine, 2002), suggesting that O-fucose is essential for all pathways mediated by Notch receptor and its ligands.

Recently, O-fucose was found in a different protein context: Thrombospondin type 1 repeats (TSRs) (Figure 1B). TSR O-fucosylation was first described on all three TSRs
Galactose  Glucose  Sialic acid  Fucose

Monosaccharide fucose to a fully elongated tetrasaccharide. (A) O-Fucose is linked to S or T within the consensus sequence C^2\text{X}_2\text{X}_3(S/T)C^3, where C^2 and C^3 are the second and third conserved C residues of an EGF-like repeat. O-Fucosylglycans can exist in the form of the monosaccharide fucose to a fully elongated tetrasaccharide. (B) O-Fucose is also linked to S or T of the consensus sequence WX_cC^1\text{X}_2\text{X}_3(S/T)C^2\text{X}_4\text{X}_5(G, where C^1 and C^2 are the first and second conserved C residues of a TSR. Both mono- and disaccharide forms of the O-fucosylglycan have been reported on TSRs (Hofsteenge et al., 2001; Luo, Nita-Lazar, et al., 2006). The enzymes responsible for the formation of glycosidic linkages are indicated. Their abbreviations are β3GlicNAcT for β1,3-N-acetylglucosaminyltransferase belonging to the Fringe family, β4GalT-I for β1,4-galactosyltransferase 1, SiaT for sialyltransferase, and β3GlcT for β1,3-glucosyltransferase.

Results

Genomic organization of bovine Pofut1 and Pofut2 genes

Homology searches against bovine EST database using human and mouse Pofut1 and Pofut2 coding sequences (CDSs) as queries revealed the existence of Pofut orthologous sequences. These bovine ESTs allowed the design of primers for cDNA isolation from bovine tissues (Table 1 in Supplementary Data). The sequencing of full-length cDNA and comparison with human and mouse genomic sequences were used to design specific primers for the determination of exon–intron boundaries of the bovine genes (Table 2 in Supplementary Data). Bovine Pofut1 consists of at least 10 exons encompassing 39.4 kb, and bovine Pofut2 consists of nine exons encompassing 13.9 kb (Figure 2). Coding exons range in size from 90 to 803 bp for Pofut1 and from 67 to 1050 bp for Pofut2. The size of introns varies from 1153 bp to at least 16,097 bp for bovine Pofut1 and from 215 bp to at least 5193 bp for bovine Pofut2. Large introns were not entirely sequenced; they were amplified by polymerase chain reaction (PCR), and their sizes were estimated by agarose gel electrophoresis and then compared with sizes deduced from the partially sequenced cow genome (www.ncbi.nlm.nih.gov/Genomes/). Exon–intron boundary sequences follow the “GT and AG” rule. The exon–intron organization of bovine Pofut1 and Pofut2 genes resembles those of human and mouse genes. However, exons 5, 9, and 10 of bovine Pofut1 have never been described before. Southern blot experiments with an exon 7 probe for Pofut1 and an exon 2 probe for Pofut2 were carried out with digested bovine genomic DNA extracted from whole blood (Figure 3). For each of the three performed digestions, only one band was detected for each gene, suggesting that only one Pofut1 gene and one Pofut2 gene are present in the bovine genome.

Detection of alternatively spliced Pofut1 and Pofut2 transcripts in bovine tissues

Reverse transcriptase (RT)–PCR analyses with specific bovine Pofut1 primers designed in exon 1 and in exon 8 or 10 (Table 2 in Supplementary Data) revealed at least five different bovine Pofut1 transcript variants (Figure 4A) named Pofut1a to Pofut1e. Pofut1a (GenBank accession
B. intron sequences in lowercase, and donor and acceptor splice sites in bold-

exons 1–4 and 6–8. It encodes the putatively active enzyme

to direct sequencing of a bacterial artificial chromosome (BAC) containing the bovine Pofut2 gene. A putative 131-

number AY344580) corresponds to the transcription of exons 1–4 and 6–8. It encodes the putatively active enzyme

In the same way, RT–PCR with specific Pofut2 primers revealed five different transcripts in bovine tissues (Pofut2a to Pofut2e) (Figure 4B). Pofut2a (DQ138294) results from the use of a splice site 90 nucleotides within exon 8 and encodes the putatively active bovine Pofut2. Indeed, it has a predicted size of 429 amino acids, consistent with the human and mouse orthologs (Luo, Koles., et al., 2006). Pofut2b (AY344582) corresponds to the transcription of all exons and would encode a protein with a predicted sequence of 459 amino acids. Pofut2c (DQ138295), Pofut2d (DQ138296), and Pofut2e (DQ138297) result from alternative splicings involving exons 4, 7, and 8. Each of these variants would encode the same putative protein of 186 amino acids because of a frame shift in exon 4 and, consequently, the occurrence of a premature stop codon in exon 5.

To determine the transcription start site for the Pofut genes and to obtain further insights into the 5' untranslated region (UTR), we performed 5' rapid amplification of cDNA end (5'-RACE) experiments, using heart mRNA. Nested PCR amplification and sequencing allowed the identification of a 68-base-long Pofut1 5' UTR. cDNA sequences of Pofut1a, Pofut1c, and Pofut1d have the same 3' UTR with a putative partial polyadenylation signal (ATAAA) at positions +1840, +1554, and +1404 bases, respectively. Pofut1b and Pofut1e possess a different 3' UTR, due to the presence of exons 9 and 10, with a putative polyadenylation signal (AATAAA) at positions +1424 and +988 bases, respectively. In summary, total Pofut1 transcript sizes range from 1009 to 1858 bases. The 5'-RACE experiment on Pofut2 was not conclusive. Instead, we proceeded to direct sequencing of a bacterial artificial chromosome (BAC) containing the bovine Pofut2 gene. A putative 131-base-long 5' UTR is predicted using the Dragon Promoter.
Bovine promoter/promoter1_5/DPF.htm). cDNA sequences revealed

bovine Fig. 4. the same 3’ +2074 bases, respectively. (AATAAA) at positions +2300, +2390, +2329, +2239, and

stop codons are indicated on each splice variant. Complete transcript sizes and predicted protein sizes are indicated on the right.

Finder online software (http://research.i2r.a-star.edu.sg/promoter/promoter1_5/DPF.htm). cDNA sequences revealed

that all Pofut2 transcript variants (Pofut2a to Pofut2e) possess

the same 3’ UTR with a putative polyadenylation signal (AATAAAA) at positions +2300, +2390, +2329, +2239, and

+2074 bases, respectively. Pofut2 transcripts are longer than those of Pofut1, with lengths ranging from 2091 to 2407 bases.

RT–PCR was also performed on 13 adult bovine tissue samples to determine the expression profiles of Pofut1 and Pofut2 complete CDS. The results of such experiments are shown for Pofut1 variants in Figure 5A. All analyzed bovine tissues present one to five different Pofut1 and Pofut2 transcript variants (Figure 5B). Pofut1a and Pofut1b were detected in all examined tissues, consistent with the widespread distribution of O-fucose and the fact that these transcripts may encode active forms of Pofuts. Pofut1b is only present in brain and longissimus thoraci muscle, and Pofut1c in liver. The shortest Pofut1 transcripts exist in six tissues for Pofut1d and in eight for Pofut1e. Pofut2b to Pofut2e were detected in all examined bovine tissues except in rectus abdominis muscle and spleen for Pofut2c. In all

adult bovine tissues tested, Pofut1 and Pofut2 genes appear to be weakly expressed. Indeed, nested PCR had to be used to detect the presence of transcript variants.

Sequence characteristics and enzyme activities of bovine O-fucosyltransferases

Protein encoded by Pofut1a (Pofut1A) possesses a potential N-terminal signal peptide (M1 to V29), two putative N-glycosylation sites (N65 and N162), the three conserved peptide motifs common to α1,2- and α1,6-fucosyltransferases (E236, W248, S308, E321, and E342, E368) (Martinez-Duncker et al., 2003), a DXD motif (E368, E369, D370) that is believed to be critical for the proper function of many glycosyltransferases (Munro and Freeman, 2000), and a KDEL-like motif for ER-resident proteins (the last four amino acids: RDEF). Pofut1B lacks the third conserved peptide motif, the DXD, and the KDEL-like motifs. Pofut1C, Pofut1D, and Pofut1E with predicted sequences of 229, 184, and 185

Fig. 4. Bovine Pofut1 and Pofut2 transcript variants. RT–PCR analyses of bovine Pofut1 (A) and Pofut2 (B) splice variants revealed five different types for each gene (Pofut1a to Pofut1e and Pofut2a to Pofut2e, respectively). Numbered black boxes represent exons. The translation start and stop codons are indicated on each splice variant. Complete transcript sizes and predicted protein sizes are indicated on the right.

Fig. 5. Expression of Pofut1 and Pofut2 splice variants in different bovine tissues. (A) Total RNAs isolated from 13 adult bovine tissues were reverse transcribed, amplified by primer pairs specific to each Pofut1 variant (see Table 1 in Supplementary Data), and subjected to agarose gel electrophoresis. Pofut1 variants are indicated on the left and bovine tissues above. β-Actin amplification was used as a control of mRNA integrity. (B) Summary of the expression of Pofut1 and Pofut2 splice variants in the 13 adult bovine tissues. The presence or absence of the variants is illustrated with the symbols + or −, respectively.
amino acids, respectively, are largely truncated and lack all the conserved peptide motifs.

To check whether bovine Pofut1a produces an active enzyme, we assessed the activity of Pofut1A using a bacterially expressed EGF-like repeat from human factor IX as described by Rampal et al. (2005). To confirm the expression of the recombinant enzyme, we performed western blot analysis using rabbit polyclonal antibodies raised against bovine Pofut1. A protein of 45 kDa, which corresponds to the predicted size of glycosylated Pofut1A, was strongly detected in extracts from COS-1 cells transfected with the Pofut1a CDS construct (Figure 6A, lane 2). Comparatively, endogenous Pofut1 was only weakly detected (Figure 6A, lane 1). Mock-transfected COS-1 cells revealed an endogenous activity of 4.5 nmol/h/mg of total proteins, whereas an activity of 53.6 nmol/h/mg was measured after transfection by the bovine Pofut1a CDS construct (Figure 6A), around 12 times higher than the control. The recombinant proteins encoded by variants Pofut1b to Pofut1e, although detected on western blots for Pofut1B and Pofut1C (data not shown), are not able to catalyze the transfer of fucose on EGF-like repeat (data not shown).

Pofut2A possesses a potential N-terminal signal peptide (M1 to Q36), three putative N-glycosylation sites (N189, N209, and N259), the three conserved peptide domains (P287–I299, K237–E340, and I369–E396), and the DXD motif (E396 R397 E398). No KDEL-like motif was found. Compared to this sequence, Pofut2B shows an insertion of 30 amino acids in the second conserved peptide motif, whereas Pofut2C, Pofut2D, and Pofut2E, all of which encode the same protein, are largely truncated and lack all the conserved peptide motifs.

Pofut2 activity was assayed using a bacterially expressed TSR from human Thrombospondin-I as acceptor substrate (Luo, Nita-Lazar, et al., 2006) (Figure 6B). The extracts of mock-transfected COS-1 cells had an endogenous activity of 12.4 nmol/h/mg. After transfection with the bovine Pofut2a CDS construct, an activity of 16.2 nmol/h/mg was measured, increased by 30% compared with that of the control. Similar results were seen when expressing recombinant mouse Pofut2 (Dlugosz, M.A. and Haltiwanger, R.S., in preparation). Using the same method, we also assayed the activities of the other putative proteins encoded by variants Pofut2b to Pofut2e. Only Pofut2A was able to cause a statistically significant increase in the transfer of fucose to TSR over endogenous activity (Figure 6B). Immunoblot analysis showed that antibovine Pofut2 antibody cross-reacts with the endogenous Pofut2 in COS-1 cells (Figure 6B, lane 1).

Fig. 6. O-Fucosyltransferase activity of bovine Pofut1 and Pofut2. The CDSs for bovine Pofut1A and Pofut2A were cloned into the pcDNA/TOPO3.1 eukaryotic expression vector. Recombinant vectors were used to transiently transfect COS-1 cells. Activities are given in nmol of fucose transferred per hour and per mg of total proteins. The expression of endogenous and recombinant proteins was assessed by immunodetection of 50 μg of total proteins with polyclonal antibodies raised against bovine Pofut1 (A, lanes 1 and 2) or Pofut2 (B, lanes 1 and 2). In each case (lane 3), 100 ng of bacterially expressed bovine Pofuts was used as positive controls. The expected size of Pofut1 expressed in Escherichia coli is 46.9 kDa because it contains 51 additional amino acids corresponding to the signal peptide and one (His)6-Tag at each end of Pofut1. In the same way, the Pofut2 expressed in E. coli contains 52 additional amino acids and its size is 52 kDa. (A) Five micrograms of total protein extracts was used to assay O-fucosyltransferase activity using a bacterially expressed EGF-like repeat from human factor IX. COS-1 cells transfected with the Pofut1a CDS construction revealed an activity of 53.6 nmol/h/mg, and mock-transfected COS-1 cells an endogenous activity of 4.5 nmol/h/mg. (B) Ten micrograms of total protein extracts was used to assay O-fucosyltransferase activity using a bacterially expressed TSR from human Thrombospondin-I. COS-1 cells transfected with the Pofut2a CDS construction revealed an activity of 16.2 nmol/h/mg, and mock-transfected COS-1 cells an endogenous activity of 12.4 nmol/h/mg. These activities are mean values of three independent experiments. Bars on bar graphs represent standard deviations.
Reactivity toward a 50-kDa protein, consistent with the predicted size of glycosylated Pofut2, increased with COS-1 cells transfected with the Pofut2a CDS construct (Figure 6B, lane 2). Among the proteins encoded by variants Pofut2b to Pofut2e, only Pofut2B was detected in transfected COS-1 cells (data not shown).

**Phylogeny of Pofut1 and Pofut2 enzymes**

Database searches allowed the selection of complete sequences (except for Aedes aegypti and Onchocerca volvulus Pofut1, and Oryzias latipes and Takifugu rubripes Pofut2) from a panel of species representing most of the animal kingdom diversity. For the two enzymes, only one copy of the gene is present in all completely sequenced genomes. Twenty-seven protein sequences corresponding to Pofut1 orthologs and 21 sequences corresponding to Pofut2 orthologs were selected and aligned. Protein sizes vary from 380 to 404 amino acids for Pofut1 and from 405 to 495 amino acids for Pofut2. The main variation resides in the N-terminal part of the enzymes, which includes the signal peptide. Protein maximum likelihood (ML) analysis with bootstrap replicates (BP) and a Bayesian phylogenetic search with posterior probabilities (PPs) were carried out.

The classical phylogenetic relationships between animal clades were recovered for both enzymes (Figure 7). Mammals, actinopterygians, deuterostomes, insects (only for Pofut1), and nematodes form monophyletic groups with BP ranging from 80 to 100% and PP of 1 (with the exception of intron 1. Therefore, the last common ancestor of all parasites of vertebrates. No Pofut1 was found in these species. Only Cryptosporidium Pofut2 were included in our analyses because Plasmodium berghei, P. falciparum, and Plasmodium vivax sequences would introduce some noise in the phylogenetic reconstruction because of their high levels of divergence. Both species of Cryptosporidium are sister group in the Pofut2 tree and present the longest branches.

**Discussion**

In this article, our work focused on the genomic organization, identification, and distribution of alternative spliced variants, enzymatic activity, and phylogeny of O-fucosyltransferases, in the context of the nearly completely sequenced genome of the animal model B. taurus.

In silico analyses revealed that the bovine Pofut1 gene is located on chromosome 13 close to pleomorphic adenoma-like gene 2 (Plagl2), underlying a conserved synteny with human (chromosome 20q11) and mouse (chromosome 2H2). The exon–intron organization of bovine Pofut1 gene is slightly different from those of human and mouse orthologous genes. Bovine Pofut1 gene is longer and contains two additional exons (exons 9 and 10) located at least 16 kb downstream the exon homolog to the last one in human and mouse genes. No homology with bovine exons 9 and 10 was detected in sequences of the large diversity of species studied in this article. Compared with mouse, bovine and human Pofut1 show an additional exon at position 5, but the sequences of both exons are not homologous. However, a homologous sequence of bovine exon 5 is found in human intron 6. Moreover, the transcription of exon 5 in both cases generates a premature stop codon, and if the translation occurs, a truncated inactive enzyme will be produced. In B. taurus, this transcript is tissue specific, restricted to the liver, whereas in Homo sapiens, the homologous transcript, named variant 2 (BC000582), was found in brain.

The active Pofut1 enzyme is encoded in B. taurus by the transcript variant Pofut1a that includes exons 1–4 and 6–8. In mammals, birds, fishes, and amphibians, the six intron positions are strictly conserved (Figure 7A). Variations observed in protein sizes only result from variation in exon 1 length. These small differences at the N-terminus part of Pofut1, in the portion corresponding to the signal peptide, explain the size distribution from 380 amino acids for Xenopus and Gallus gallus to 395 amino acids for Rattus norvegicus, Danio rerio, and O. latipes. This six-intron organization is also recovered in the sea urchin S. purpuratus (data not shown) and in the nematode Caenorhabditis. In the first, compared with vertebrates, exon size variations were observed, but nearly all intron positions are conserved with only a small shift in the position of intron 3. In the second, only positions of introns 1 and 5 are strictly conserved. The same situation is present in the honeybee Apis mellifera, although the sequence is 3′ partial (so not included in our analysis). However, exon organization is different among insect species. Drosophila shows two exons, whereas B. mori presents four, but all insect Pofut1 genes conserve the position of intron 1. Therefore, the last common ancestor of all analyzed bilaterian species would already possess a Pofut1 gene with at least two exons and one intron at position 1. A broader panel, especially of protostome Pofut1 genes, is needed to settle this conclusively and to demonstrate whether intron 5 was also present in the ancestral Pofut1
that 26,500 genes). A parsimonious explanation is to consider approximately one gene per 100 kb in the human (around 16,000 genes) compared to
in Ciona intestinalis (around 16,000 genes) compared to approximately one gene per 100 kb in the human (around 26,500 genes). A parsimonious explanation is to consider that the last common ancestor of bilaterians already possesses at least three coding exons with introns 2 and 3 in conserved positions. However, the Bayesian tree does not support this group (PP < 0.5). If we admit that protostomes are a monophyletic group (Philippe, Lartillot, et al., 2005) and are artifactually split in our tree, then the most parsimonious hypothesis is to consider that the last common ancestor of deuterostomes already possesses at least the three-exon architecture. The introns were lost in the insect lineage and for Ciona, Plasmodium, and Cryptosporidium for Pofut2 genes. Based on the ML tree, nematodes and deuterostomes would share a common ancestor containing in its genome a Pofut2 gene with at least three coding exons with introns 2 and 3 in conserved positions. However, the Bayesian tree does not support this group (PP < 0.5). If we admit that protostomes are a monophyletic group (Philippe, Lartillot, et al., 2005) and are artifactually split in our tree, then the most parsimonious hypothesis is to consider that the last common ancestor of bilaterians already possesses at least the three-exon architecture. The introns were lost in the insect lineage and for Ciona, as found for Ciona Pofut1.

Interestingly, Pofut2 genes exist in two genera of apicomplexan parasites, Cryptosporidium and Plasmodium, where they correspond to monoexonic and bieoxonic genes, respectively. This intron position is not shared by any other Pofut2
genes sequenced to date. As Cryptosporidium is an early emerging genus among the phylum Apicomplexa (Zhu et al., 2000; Bankier et al., 2003), it could be hypothesized that the intron appeared in the common ancestor of Plasmodium species, which emerged later. Another plausible explanation comes from the observation that the number of predicted introns in Cryptosporidium (<10% of genes) is lower than that in Plasmodium (>50%), which had to be linked with the loss of splicosomal machinery components in Cryptosporidium (Templeton et al., 2004). Therefore, the reduction in the number of introns would be then considered as a lineage-specific evolutionary phenomenon. The Pofut2 gene in apicomplexans was probably acquired by lateral gene transfer from one host before the speciation of Cryptosporidium and Plasmodium. However, at present, there is no evidence of Pofut2 in the genome of other apicomplexans such as Toxoplasma gondii. The presence of a potential O-fucosyltransferase could be related to the existence of numerous proteins containing conserved adhesion domains, such as EGF-like repeats and TSRs (Templeton et al., 2004).

Tissue distributions of human and mouse Pofut1 transcripts have only been studied by northern blots (Wang et al., 2001; Shi and Stanley, 2003). The authors showed that many Pofut1 splice variants exist in various tissues, but only the transcript encoding the putative active O-fucosyltransferase was described. Human variant 1 (exons 1–4 and 6–8) and the mouse transcript, encoding active forms of Pofut1 consisting of 393 and 388 amino acids, respectively, have been detected in all examined tissues (including heart, brain, liver, skeletal muscle, and kidney), consistent with the widespread localization of O-fucose (Wang et al., 1996, 1998). In mice, the Pofut1 gene is increasingly expressed during embryonic development (Shi and Stanley, 2003). In humans, the transcript is 5.2 kb long (Wang et al., 2001), comparable to the estimated size in mice (Shi and Stanley, 2003). Moreover, the transcript contains an extensive 3′ UTR (around 4 kb), typical of many glycosyltransferases (Breton and Imberty, 1999). In this study, we demonstrated using RT–PCR that bovine tissues contain at least one to five different transcript variants. The longest bovine transcript, Pofut1a, is 1858 bases long with a 3′ UTR of only 631 bases, perhaps revealing a different Pofut1 gene expression control mechanism in cattle. Pofut1a encodes the active enzyme and is detected in the 13
examined adult tissues. Interestingly, the shortest Pofut1 variants (Pofut1b to Pofut1e), which result from different alternative splicings, were not detected in all tissues. They could be specific to B. taurus because none of these variants are found in other animal EST databases.

Five bovine Pofut2 transcripts (Pofut2a to Pofut2e) were identified and are present in nearly all examined adult bovine tissues. Among them, only bovine Pofut2a transcript encodes the active enzyme and is clearly orthologous to mouse transcript (Luo, Koles, et al., 2006). No accurate experiment was conducted on the tissue distribution of Pofut2 variants in other species. However, searches in animal EST databases failed to detect variants homologous to Pofut2b to Pofut2e. Only two human ESTs (BP318733 and DB158060) from pericardium and thymus would encode a putative protein of 186 amino acids, similar to bovine Pofut2C, Pofut2D, and Pofut2E. Database searches also show that the majority of proteins containing TSFs possess the consensus TSR O-fucosylation site (Hofsteenge et al., 2001), indicating that this modification could be common. This is in agreement with the large distribution of Pofut2a transcript in bovine tissues and with the presence of the corresponding gene in a wide variety of organisms.

Bovine Pofut1 has a signal peptide and a KDEL-type sequence (RDEF). Therefore, as demonstrated for the human enzyme (Luo and Haltiwanger, 2005), the bovine Pofut1 could be a soluble protein located in the lumen of the ER. Drosophila Ofut2 also appears to be localized to the ER (Luo, Koles, et al., 2006) even though the enzyme does not possess an ER retention signal. An ER localization is also predicted for Pofut1 and Pofut2 of all the animal species studied in this article. Thus, as Pofut1 and Pofut2 only act on properly folded domains in the ER, O-fucosyltransferases may constitute a new class of quality-control proteins (Luo and Haltiwanger, 2005; Luo, Koles, et al., 2006).

Materials and Methods

Identification of bovine Pofut1 and Pofut2 cDNA sequences and genomic structures

Searches of bovine EST databases were performed using the Basic Local Alignment Tool (BLAST) at the National Centre for Biotechnology Information (NCBI, at www.ncbi.nlm.nih.gov/BLAST/) with human and mouse Pofut CDS as queries. Two bovine Pofut1 ESTs were found—BF655050 from a pooled tissue database (marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle) and AW415368 from another pooled tissue database (lymph node, ovary, fat, hypothalamus, and pituitary). One Pofut2 EST, accession number AW669493, originating from the latter database, was also identified. The full-length ESTs, cloned in the pcMV SPORT6 vector, were obtained from Institut National de la Recherche Agronomique (INRA, Centre de Ressources Biologiques en Génomique des Animaux Domestiques et d’Intérêt Economique, Jouy-en-Josas, France) and sequenced. Comparison with human and mouse genomic organizations using the Spidey mRNA-to-genomic alignment program (www.ncbi.nlm.nih.gov/spidey/) allowed the design of specific primers in the exon borders flanking intronic regions (Table 2 in Supplementary Data). These primers were used for direct sequencing of genomic DNA extracted from bovine Charolais whole blood (QIAmp Blood Kit, Qiagen Inc., Hilden, Germany). Exons and introns were amplified by PCR, allowing the determination of bovine Pofut gene structures. All exons and small introns were completely sequenced. When possible, the sizes of large introns were estimated, after PCR amplification, by agarose gel electrophoresis. The increasing effort in the cow genome sequencing and assembling by the international consortium allowed us to confirm some sequences and intronic sizes, but some information is still missing for these two genes. One bovine BAC, clone 0669H8, containing the appropriate part of the bovine Pofut2 gene was used for direct sequencing (Jeon et al., 2001) of the Pofut2 5’ end, with the primer BosPofut2/5’ end (Table 2 in Supplementary Data).

Southern blot analysis of bovine genomic DNA

Bovine genomic DNA was prepared from blood sample (QIAmp Blood Kit). Ten micrograms of the DNA was subjected to digestion with EcoRI, HindIII, or NdeI endonucleases, and the fragments were separated in 0.8% (w/v) agarose gel. DNA was depurinized for 20 min with 0.25 N HCl, denatured for 30 min with 0.4 N NaOH, and transferred onto a Hybond-N1 membrane (Amersham Pharmacia Biotech, Europe GmbH, Orsay, France). Hybridizations were carried out with an exon 7 probe for Pofut1 and an exon 2 probe for Pofut2, generated by PCR amplification of genomic DNA using specific primers: 5′-ACGCGTGTCAGCACGGGAAG-3′ forward and 5′-AGAATCGGTCGGCGGAGCA-3′ reverse primers for Pofut1, and 5′-CCTTTCTGTACGTAGTCAATCC-3′ forward and 5′-GGATGAACTCTGACGCTTCTTCT-3′ reverse primers for Pofut2. Twenty-five nanograms of probes was labeled with [α-32P]dCTP using a random priming kit (Invitrogen, Carlsbad, CA). Labeling was performed overnight at 65°C in a buffer containing 10% (w/v) dextran sulfate, 1% (w/v) sodium dodecyl sulfate (SDS), 0.5 M NaCl, and 100 μg of sheared salmon sperm DNA. Blots were washed two times for 15 min each at 65°C with 2× sodium chloride–sodium citrate buffer (SSC), two times with 1× SSC, and one time with 1× SSC/0.1% SDS, then analyzed after exposure to X-ray films (Kodak, Kyoto, Japan) for 3 days at –80°C.

Total RNA extraction and RT–PCR analysis

Bovine tissues had been sampled on a 15-month-old Charolais at the INRA Centre de Clermont-Ferrand Theix (France). Total RNAs from various bovine tissues were isolated with FastRNA Pro Green Kit (Q-BioGene, Illkirch, France) according to the manufacturer’s instructions, and first-strand cDNA was synthesized from 2 μg of total RNA templates with the SuperScript First Strand Synthesis System for RT–PCR (Invitrogen). Potential transcripts were investigated by designing primers from either side of the CDS (Table 2 in Supplementary Data). The 50-μL PCR reactions were run with 2-μL RT products, 0.3 pm sense and antisense primers (MWG, High Point, NC), 0.2 mM dNTP, 2 mM MgCl2, 1× PCR buffer, and 0.5 U of Taq polymerase (Interchim,
Montluçon, France). Amplification of the target cDNA was performed in a Primus 96 plus thermocycler (MWG). After 3 min denaturation at 96°C, 35 PCR cycles were carried out as follows: 30 s at 94°C, 45 s at 60°C, 2 min at 72°C, and a final extension of 5 min at 72°C. Reaction products were analyzed by electrophoresis on 1% agarose gel and compared with 1Kb+ DNA Ladder (Invitrogen). The RT–PCR products were cloned into the pGEM-T-Easy plasmid (Promega, Madison, WI) for entire sequencing using T7 and SP6+ vector primers and internal primers. A dye-labeling chemistry kit (PRISMReady Reaction Ampli Taq FS, Applied Biosystems, Norwalk, CT) and the ABI Prism 310 Genetic Analyzer (Applied Biosystems, Norwalk, CT) were used for sequencing.

**5′ Rapid amplification of cDNA end (5′-RACE)**

To obtain 5′-cDNA ends, we performed RACE experiments on 1 μg of total RNA extracted from bovine heart, using the SMART RACE cDNA Amplification kit (Ozyme, Montigny le Bretonneux, France) according to the manufacturer’s instructions. 5′ UTR of Pofut1 transcript was amplified by two rounds of PCR combining each time-specific and adaptor primers: Pofut1–5′RACE (5′-ATCCACTG-GTCTTTTGAGGAGGACACTG-3′)/universal primer mix (UPM) and Pofut1–5′RACE (5′-TCAGCTGCGGGTGAAATCTG-3′)/nested universal primer (NUP). First- and second-round PCRs were carried out in a 50-μL reaction volume containing 0.3 pM sense and antisense primers, 0.2 mM dNTP, 2 mM MgCl2, 1× PCR buffer, and 0.5 U of Taq polymerase in the following cycling conditions: initial denaturation at 94°C for 3 min followed by 35 cycles (94°C for 30 s, 65°C for 45 s, and 72°C for 2 min) and one last elongation step (72°C for 10 min). PCR products were cloned into the pCR2.1 vector (Invitrogen) and sequenced.

**Expression of bovine O-fucosyltransferases in COS-1 cells**

Pofut1 and Pofut2 variants were cloned into the pcDNA/TOPO3.1 expression vector (Invitrogen). Highly pure recombinant plasmids were obtained by anion-exchange chromatography (plasmid midi Kit, Qiagen, Les Ulis, France). They were used to transiently transfect COS-1 cells using Fugene 6 Transfection Reagent (Roche, Mannheim, Germany). They were used to transiently transfected COS-1 cells using Fugene 6 Transfection Reagent (Roche, Mannheim, Germany). After 48 h of transfection, cells were harvested and lysed in a lysis buffer (1% [v/v] Triton X-100, 10 mM sodium cacodylate (pH 6), 20% [v/v] glycerol and 1 mM dithiothreitol [DTT]) for 2 h at 4°C. Protein concentration was determined with bovine serum albumin (Bio-Rad, Marne-la-Coquette, France) as a standard (Bradford, 1976).

**Western blot analyses**

Polyclonal antibodies were obtained by immunization of rabbits with recombinant bovine Pofut1 or Pofut2. Enzymes were produced in *Escherichia coli* BL21(DE3) cytoplasm associated with an NH2-(His)6-Tag and were purified by affinity chromatography on Ni-NTA-agarose (Qiagen, Les Ulis, France). Subsequently, IgG was purified from serum using protein G (Agro-Bio, La Ferté St. Aubin, France). Western blot analyses were performed mainly as described (Dupuy et al., 2002) and bacterially expressed bovine Pofuts were used as controls.

**O-Fucosyltransferase assay**

O-Fucosyltransferase assays were performed essentially as described (Wang and Spellman, 1998; Luo, Nita-Lazar, et al., 2006). The reaction mixture (50 μL) contains 0.1 M imidazole–HCl (pH 7.0), 50 mM MnCl2 (only for O-fucosyltransferase 1 assay), 0.1 mM of GDP-[3H]fucose (4000–8000 cpm/nmol, Perkin-Elmer, Norwalk, CT), 20 μM of acceptor substrate (recombinant human factor IX EGF-like repeat for Pofut1 or recombinant human Thrombospondin-1 TSR3 for Pofut2) and 5 or 10 μg of total COS-1 cells protein extract for Pofut1 or Pofut2, respectively. The mixture was incubated at 37°C for 7.5 min for Pofut1 and 3 h for Pofut2. The reaction was stopped on ice and then diluted with 950 μL of 0.25 M ethylenediamine tetraacetic acid (EDTA), pH 8.0. Samples were loaded onto a Sep-Pak C18 cartridge (Waters, Saint-Quentin, France). The cartridge was washed with 10 mL of H2O, and the protein was then eluted directly into scintillation vial with 3 mL of 80% ace- tonitrile containing 0.05% trifluoroacetic acid. Ten milliliters of Biodegradable Counting Scintillant (Amersham Pharmacia Biotech) was added, and radioactivity was counted using a liquid scintillation beta counter (liquid scintillation analyzer, Tri-Carb-2100TR, Packard, Meriden, CT).

**Phylogenetic analyses**

Pofut1 and Pofut2 orthologs were retrieved from various databases using BLAST. Their accession numbers in Entrez protein database are *A. aegypti* Pofut1 ABA29465, *Anopheles gambiae* Pofut1 ABA29466 and Pofut2 ABA29475, *B. mori* Pofut1 ABA29467 and Pofut2 ABA29474, *B. taurus* Pofut1 AAQ02332 and Pofut2 CAE02608, *Canis familiaris* Pofut1 ABA29461, *Caenorhabditis briggsae* Pofut1 ABA29470 and Pofut2 CAH03732, *C. elegans* Pofut1 ABA29469 and Pofut2 Q8WR51, *C. intestinalis* Pofut1 AK112708 and Pofut2 CAE02609, *Ciona savignyi* Pofut1 CAH03713 and Pofut2 ABA29472, *Cryptosporidium hominis* Pofut2 EAL37907, *Cryptosporidium parvum* Pofut2 EAK88374, *D. rerio* Pofut1 NP_991280 and Pofut2 ABA29477, *Drosophila melano- gaster* Ofut1 Q9V6X7 and Ofut2 AK77300, *Drosophila pseudoobscura* Ofut1 DQ139947 and Ofut2 ABA29473, *Drosophila yakuba* Ofut1 CAH40384 and Ofut2 CAH41976, *G. gallus* Pofut1 NP_989758 and Pofut2 XP_421892, *H. sapiens* Pofut1 NP_056167 and POFUT2 NP_598368, *M. musculus* Pofut1 NP_553671 and Pofut2 NP_084538, *O. volvulus* Pofut1 CAH40836, *O. latipes* Pofut1 ABA29462 and Pofut2 ABA29476, *Pan troglodytes* Pofut1 CAH03712 and Pofut2 NP_001008983, *Pongo pygmaeus* Pofut1 CAH91412, *R. norvegicus* Pofut1 NP_001002278 and Pofut2 XP_228073, *S. purpuratus* Pofut1 ABA29464, *S. scrofa* Pofut1 ABA29460, *T. rubripes* Pofut1 CAE54305 and Pofut2 A1781759, *Tetraodon nigroviridis* Pofut1 ABA29463, *Xenopus laevis* Pofut1 CAD55833, *Xenopus tropicalis* Pofut1 CAH03710 and Pofut2 ABA29471. When needed, complete CDSs were reconstructed by EST assem- blages or searches in specific genomic databases. Alignments were performed using the PC package MUST 2000 developed by Hervé Philippe from the University of Montreal and downloadable at www.isem.univ-montp2.fr/PPP/PM/RES/Info/@Softwares.php# MUST2000. They are available from the authors upon request. Positions that could...
not be confidently aligned due to indels were removed from analyses. We used 302 positions corresponding to 227 parsimony informative characters for Pofut1 comparison and 382 positions (330 parsimony informative characters) for Pofut2. Owing to this selection, H. sapiens and P. troglodytes Pofut1 sequences are identical. Phylogenetic trees were constructed using ML method implemented in the PROTML version 2.3 software (Adachi and Hasegawa, 1996). ML trees were built by the quick-add Operational Taxonomic Units (OTUs) search, the Jones–Taylor–Thornton matrix (JTT)-f model of amino acid substitution, and retaining the 100 top-ranking trees. Bootstrap proportions were calculated by the resampling estimated log-likelihood (RELL) method (Kishino et al., 1990) upon all top-ranking trees. A Bayesian analysis based on the posterior probabilities of phylogenetic trees was performed using the MrBayes (v.3.0b4) program (Huelsenbeck and Ronquist, 2001). A general time-reversible model (Rodriguez et al., 1990) with a Γ distribution was applied to take into account among site rate variation. Four Markov chains were used with 500,000 generations, and a tree sampling was taken every 100 generations. All trees sampled before the stationary distribution were discarded as burn-in.

Supplementary Data
Supplementary data are available at Glycobio online (http://glycob.oxfordjournals.org/).

Acknowledgments
The authors thank the “Centre de Ressources Biologiques en Génomique des Animaux Domestiques et d’Intérêt Économique” (CRB GADIE, INRA, Jouy-en-Josas, France) for providing bovine ESTs and BACs, the slaughter house of the INRA centre of Clermont-Ferrand-Theix (France) for animal samples, and M. P. Laforêt for her technical assistance in cell culture. This work has been performed in the frame of the French network “G3.” It was supported by “Conseil régional du Limousin,” an NIH (GM61126) grant to R.S.H., and a “Ministère de la Recherche et de la Technologie” doctoral fellowship to C.L.

Conflict of interest statement
None declared.

Abbreviations
BAC, bacterial artificial chromosome; BP, bootstrap proportion; CAZy, carbohydrate-active enzymes; CDS, coding sequence; EGF-like, epidermal growth factor-like; ER, endoplasmic reticulum; EST, expressed sequence tag; ML, maximum likelihood; PCR, polymerase chain reaction; Pofut1, protein O-fucosyltransferase 1; Pofut2, protein O-fucosyltransferase 2; PP, posterior probability; RACE, rapid amplification of cDNA end; RT, reverse transcriptase; SSC, sodium chloride–sodium citrate buffer; TSR, Thrombospondin type 1 repeat; UTR, untranslated region.

References


