Human Molecular Genetics, 2005, Vol. 14, No. 24 **3945–3953**
doi:10.1093/hmg/ddi418
Advance Access published on November 21, 2005

T_{-13910} DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity *in vitro*

Rikke H. Lewinsky[1,†], Tine G.K. Jensen[1,†], Jette Møller[1], Allan Stensballe[2], Jørgen Olsen[1] and Jesper T. Troelsen[1,*]

[1]Department of Medical Biochemistry and Genetics, Panum Institute, University of Copenhagen, Building 6.4., Blegdamsvej 3, DK-2200 Copenhagen N, Denmark and [2]Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

Received September 12, 2005; Revised and Accepted November 1, 2005

Two phenotypes exist in the human population with regard to expression of lactase in adults. Lactase non-persistence (adult-type hypolactasia and lactose intolerance) is characterized by a decline in the expression of lactase-phlorizin hydrolase (LPH) after weaning. In contrast, lactase-persistent individuals have a high LPH throughout their lifespan. Lactase persistence and non-persistence are associated with a T/C polymorphism at position −13 910 upstream the lactase gene. A nuclear factor binds more strongly to the $T_{-13\,910}$ variant associated with lactase persistence than the $C_{-13\,910}$ variant associated with lactase non-persistence. Oct-1 and glyceraldehyde-3-phosphate dehydrogenase were co-purified by DNA affinity purification using the sequence of the $T_{-13\,910}$ variant. Supershift analyses show that Oct-1 binds directly to the $T_{-13\,910}$ variant, and we suggest that GAPDH is co-purified due to interactions with Oct-1. Expression of Oct-1 stimulates reporter gene expression from the T and the $C_{-13\,910}$ variant/LPH promoter constructs only when it is co-expressed with HNF1α. Binding sites for other intestinal transcription factors (GATA-6, HNF4α, Fox and Cdx-2) were identified in the region of the −13 910 T/C polymorphism. Three of these sites are required for the enhancer activity of the −13 910 region. The data suggest that the binding of Oct-1 to the $T_{-13\,910}$ variant directs increased lactase promoter activity and this might provide an explanation for the lactase persistence phenotype in the human population.

INTRODUCTION

Lactase non-persistence is characterized by a low expression of lactase-phlorizin hydrolase (LPH) in adulthood. Most adults in the human population are lactase non-persistent, which in many cases result in an inability to digest diary products containing lactose given the symptoms of lactose intolerance. In contrast, lactase-persistent individuals have a high lactase expression throughout their lives and are lactose tolerant (1). The lactase-persistent phenotype is frequently found among Northern Europeans and their descendants and also in some populations in Africa. Lactase non-persistence is caused by a transcriptional down-regulation of the lactase-phlorizin hydrolase gene (*LCT*) during childhood. A similar down-regulation after weaning is seen in all other mammals investigated (1). Lactase persistence is a result of bypassing this regulation. Lactase persistence is inherited in an autosomal dominant manner, and it has been demonstrated that lactase persistence is caused by *cis*-acting elements located closely to the *LCT* gene (2). The lactase-persistent phenotype has been estimated to be only 5000–10 000 years old, consistent with the advantage an adult has to exploit milk as a nutritional source, after the setting up of dairy farming (3).

Identification of a T/C polymorphism at position −13 910 upstream the *LCT* gene associated with lactase persistence/non-persistence has initiated the investigation of the molecular mechanisms controlling developmental regulation of the LPH

*To whom correspondence should be addressed. Tel: +45 35327796; Fax: +45 35367980; Email: troelsen@imbg.ku.dk
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

expression. The $T_{-13\,910}$ variant has been reported to be 100% associated with lactase persistence in the European populations (4), but another yet unidentified mutation seems to exist in some lactase-persistent populations in Africa (5). The $C/T_{-13\,910}$ single nucleotide polymorphism (SNP) is located in intron 13 of the upstream minichromosome maintenance-6 gene (*MCM6*). Analyses of the gene-regulatory capacity of the −13 910 region have demonstrated a strong transcriptional enhancer activity of the region (6,7). The $T_{-13\,910}$ variant is a more effective enhancer of the LPH promoter activity than the $C_{-13\,910}$ variant, especially when analysed in differentiated intestinal cells. A nuclear factor from both intestinal and HeLa nuclear extracts binds strongly to the $T_{-13\,910}$ variant and much more weakly to the $C_{-13\,910}$ variant. Thus, position −13 910 is a part of a transcription factor binding site that is important for the enhancer activity of the region. The mutation from $C_{-13\,910}$ to $T_{-13\,910}$ creates a higher affinity for the nuclear factor. As the $T_{-13\,910}$ variant has increased enhancer activity, we have suggested that the life-long activation of the *LPH* gene in lactase-persistent individuals is caused by an increased recruitment of transcriptional activators to the $T_{-13\,910}$ variant that prevent the postweaning decline of LPH (1,6).

The first 150 bp of proximal promoter of the *LPH* gene is highly conserved among human, pig and rat (1). The transcription factors binding this conserved promoter region have been extensively studied. Binding of Cdx-2, HNF1α and GATA factors to the LPH promoter is required for promoter activity (8–19). HNF1α interacts directly with both Cdx-2 and GATA-4, and it has been shown that interactions between HNF1α and Cdx-2 and also HNF1α and GATA factors synergistically activate LPH promoter activity (8,12). In pig and rat, it is clear that additional upstream gene-regulatory regions are necessary for a high intestinal-specific LPH promoter activity (20–22).

In the current study, we have used a DNA affinity purification strategy to identify GAPDH and Oct-1 as $T_{-13\,910}$ variant interacting factors. Further analyses of the region surrounding $T_{-13\,910}$ variant reveal binding of several transcription factors expressed in the small intestinal epithelium. Binding of these factors is essential for the enhancer activity. These results indicate that lactase persistence in humans is a result of increased Oct-1-mediated enhancer activity that prevents the normal postweaning down-regulation of the lactase expression seen in other mammals. Although our results do not irrevocably demonstrate that the $T_{-13\,910}$ DNA variant is causal *in vivo* for lactase persistence, it clearly shows that the $T_{-13\,910}$ DNA variant is responsible for the increased gene-regulatory activity affecting LPH promoter activity.

RESULTS

Purification and identification of the $T_{-13\,910}$-binding factor

A DNA affinity purification strategy was applied to identify the nuclear factor binding to the $T_{-13\,910}$ variant. The purification resulted in the isolation of one prominent protein with a molecular weight of ∼35 kDa and some minor bands in the range of 60–100 kDa (Fig. 1A). The prominent band was in-gel digested by trypsin, and the protein was unambiguously identified using nanoflow LC-MS/MS mass spectrometry (23,24) to be glyceraldehyde-3-phosphate dehydrogenase (GAPDH). GAPDH exists in two forms: as a cytosolic tetrameric glycolytic enzyme and as a monomeric non-glycolytic nuclear form (25). The nuclear form of GAPDH has recently been shown to be a component of the OCA-S coactivator complex. OCA-S is essential for Oct-1-activated histone H2B transcription, and GAPDH has been shown to interact directly with Oct-1 (26). As supershift experiments showed that GAPDH does not directly interact with the $T_{-13\,910}$ variant (data not shown), we investigated whether the nuclear factor binding the $T_{-13\,910}$ variant could be Oct-1. A search for Oct-1 consensus sequence close to position −13 910 (Fig. 1B) using TRANSFAC® Professional 9.1 database revealed a non-classical Oct-1 binding sequence [Transfac accession no. M00137 (27)] located 5′ to position −13 910. The analysis also revealed that the matrix match score of $T_{-13\,910}$ variant is slightly higher than that of the $C_{-13\,910}$ variant, indicating that Oct-1 has a higher affinity to the $T_{-13\,910}$ variant than to the $C_{-13\,910}$ variant. As this correlates to our gel shift results (Fig. 1D), we decided to repeat the DNA affinity purification using nuclear extracts prepared from the intestinal epithelial cell line Caco-2 and to analyse for the presence of Oct-1 and GAPDH in the fractions by western blot analysis. The Fast-Q purification step was omitted to minimize the purification time and thereby degradation of the nuclear proteins. As demonstrated in Figure 1C, both Oct-1 and GAPDH can indeed be co-purified by the ability to bind the $T_{-13\,910}$ sequence.

Gel shift assays were used to further investigate the Oct-1 binding to the $T/C_{-13\,910}$ variants (Fig. 1D). The $T_{-13\,910}$ variant bound stronger to the nuclear factor than the $C_{-13\,910}$ variant. The formed complex was efficiently competed by excess of either a classical Oct-1 motif or a non-classical Oct-1 motif (M00137). Addition of anti-Oct-1 antibody resulted in a supershift of the complex. Furthermore, a gel shift assay using probe containing the classical Oct-1 motif resulted in a complex with the same mobility, and a similar supershift with anti-Oct-1 was seen (Fig. 1D).

Identification of multiple binding sites for intestinal expressed transcription factors in the −13 910 enhancer

A DNase footprint analysis and supershift analyses of the region surrounding the $T_{-13\,910}$ variant were performed to identify and analyse *cis*-elements surrounding the −13 910 SNP (Fig. 2A). A footprint was detected at position −13 909 to −13 934, including the Oct-1 binding site and a GATA-6 binding motif. Footprints from −13 857 to −13 817 span an HNF4α motif and a Fox/HNF3α motif. A footprint spans a putative Cdx-2 site at position −14 022 to −14 032. GATA-6, Fox, Cdx-2 and HNF4α are all well-known transcription factors important for intestinal gene expression. Cdx-2, Fox and GATA-6 have been shown to regulate LPH promoter activity by binding to the proximal promoter region of the *LCT* gene (11,14,17,18).

The protein/DNA interactions in the −13 910 enhancer region were further characterized by supershift analyses (Fig. 2B–D). A probe covering the footprint located at the −13 910 position was analysed for protein binding, using

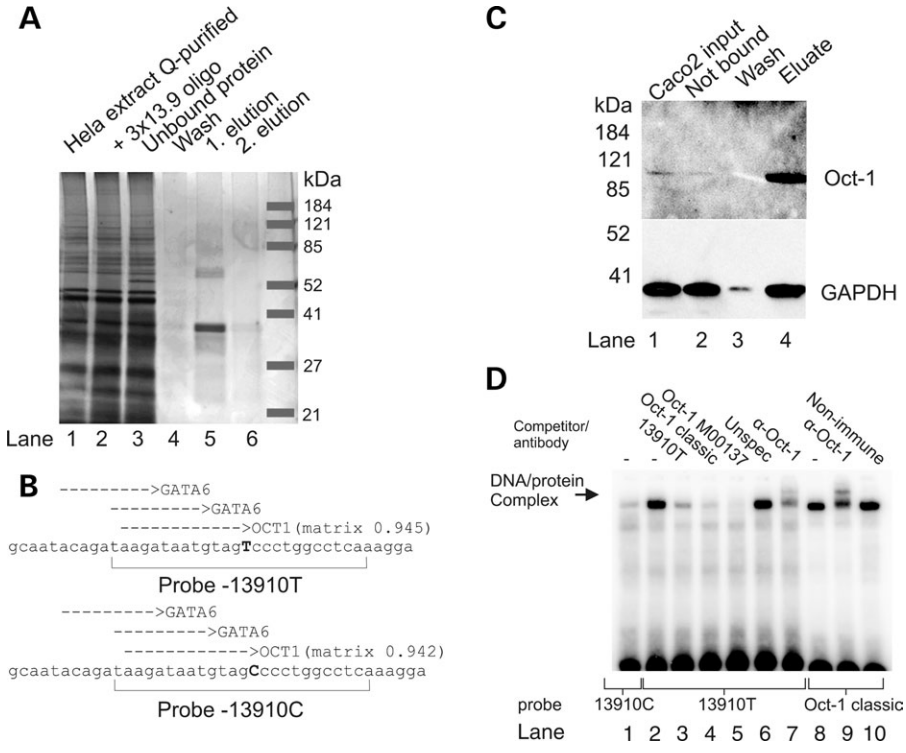

Figure 1. Oct-1 interacts with $T_{-13\,910}$ DNA sequence. (**A**) Purification of proteins interacting with the $T_{-13\,910}$ variant. HeLa-S cells were used to purify the $T_{-13\,910}$ binding activity. An ion exchange purification step (Q-fast flow column) was used to remove unspecific DNA-binding activity from the extract. The $T_{-13\,910}$ binding activity was subsequently purified by DNA affinity purification using magnetic beads coupled with the $T_{-13\,910}$ binding motif. Fractions obtained during the purification were analysed by SDS–PAGE, and the bands were visualized by silver staining. The following fractions were analysed: Q-purified HeLa-S extract (lane 1), Q-purified HeLa-S extract with biotinylated $T_{-13\,910}$ oligonucleotides (lane 2), unbound proteins after the removal of the biotinylated $T_{-13\,910}$ oligonucleotides with streptavidin-coated magnetic beads (lane 3); the $T_{-13\,910}$ magnetic beads were washed to remove unspecific bound protein (lane 4). Proteins bound to the $T_{-13\,910}$ magnetic beads were eluted twice with a high-salt buffer (lanes 5 and 6). (**B**) Identification of an Oct-1 binding site at position −13 910. A Match search at the Transfac Database (www.biobase.de) was used to identify potential transcription factor binding sites for the $T_{-13\,910}$ and the $C_{-13\,910}$ variants. The Matrix match scores of the Oct-1 motifs for the $T_{-13\,910}$ and $C_{-13\,910}$ variants are shown. (**C**) Western blot analysis of the presence of Oct-1 in the $T_{-13\,910}$ DNA affinity purification. Oct-1 can be purified from Caco-2 nuclear extract using $T_{-13\,910}$ DNA affinity purification. Lane 1, Caco-2 nuclear extract used as input in the purification; lane 2, unbound Oct-1; lane 3, wash fraction; lane 4, elution of bound Oct-1. (**D**) The canonical Oct-1 motif and the M00127 Oct-1 motif compete for binding to $T_{-13\,910}$, and the $T_{-13\,910}$/protein can be supershifted with Oct-1 antibody. Caco-2 nuclear extract was used to perform gel shift assays and supershift analysis using $C_{-13\,910}$ probe (lane 1), $T_{-13\,910}$ probe (lanes 2–7) and a canonical Oct-1 probe (lanes 8–10). Competition by unlabelled oligonucleotides was used to demonstrate specific Oct-1 binding [lane 3, $T_{-13\,910}$ and lane 4, Oct-1 classic (canonical Oct-1 site)]; lane 5, Oct-1 M00127 (alternative Oct-1 site); lane 6, oligonucleotide with unrelated sequence. Antibody against Oct-1 was shown in lanes 7 and 9. Non-immune serum was added in lane 10.

nuclear extracts from differentiated Caco-2 cells (Fig. 2C). Three specific complexes could be detected (complexes I–III). Complex I could be supershifted with an anti-Oct-1 antibody and was competed with oligonucleotides containing a classical Oct-1 site or the non-classical motif related to the $T_{-13\,910}$ sequence. Complexes II and III were supershifted with GATA-6 antibody and were not efficiently competed with the Oct-1 oligonucleotides. This indicates that only Oct-1 is present in complex I, whereas GATA-6 protein is present in complexes II and III. As $T_{-13\,910}$/Oct/GATA probe contains two GATA binding sites, it is therefore possible that complexes II and III are complexes with either one or two GATA proteins. However, complex IV supershifts with both Oct-1 and GATA-6 antibodies, showing that Oct-1 and GATA-6 are able to bind the probe at the same time and the binding is not mutually exclusive. The protein/DNA complex marked 'unspec' was not affected by the addition of unlabelled oligonucleotides with either specific or unspecific (i.e. unrelated) sequences. Thus, formation of this complex was considered to be a result of unspecific protein binding.

Supershift analyses demonstrated that HNF4α binds to the region between −13 860 and −13 825 (Fig. 2B) and Cdx-2 binds to the region −14 045 to −140 101 (Fig. 2D). Competition experiments demonstrated specific interaction of nuclear factors with the Fox/HNF3α motif. This complex could not be convincingly supershifted with an anti-HNF3α antibody, indicating that another fork head transcription factor might interact with this sequence (Fig. 2B).

Functional analysis of the −13 910 enhancer activity

A functional analysis of the $T_{-13\,910}$ and $C_{-13\,910}$ enhancer activity demonstrated, as previously shown (6), that the $T_{-13\,910}$ variant is a more powerful enhancer than the $C_{-13\,910}$ variant (Fig. 3A). The $T_{-13\,910}$ enhancer stimulates the LPH promoter-driven reporter gene expression 9-fold when compared with the LPH promoter alone (Fig. 3B). An enhancer/promoter construct was produced containing mutations in the Oct-1 site in the $T_{-13\,910}$ enhancer to address the gene-regulatory importance of the Oct-1 binding for the

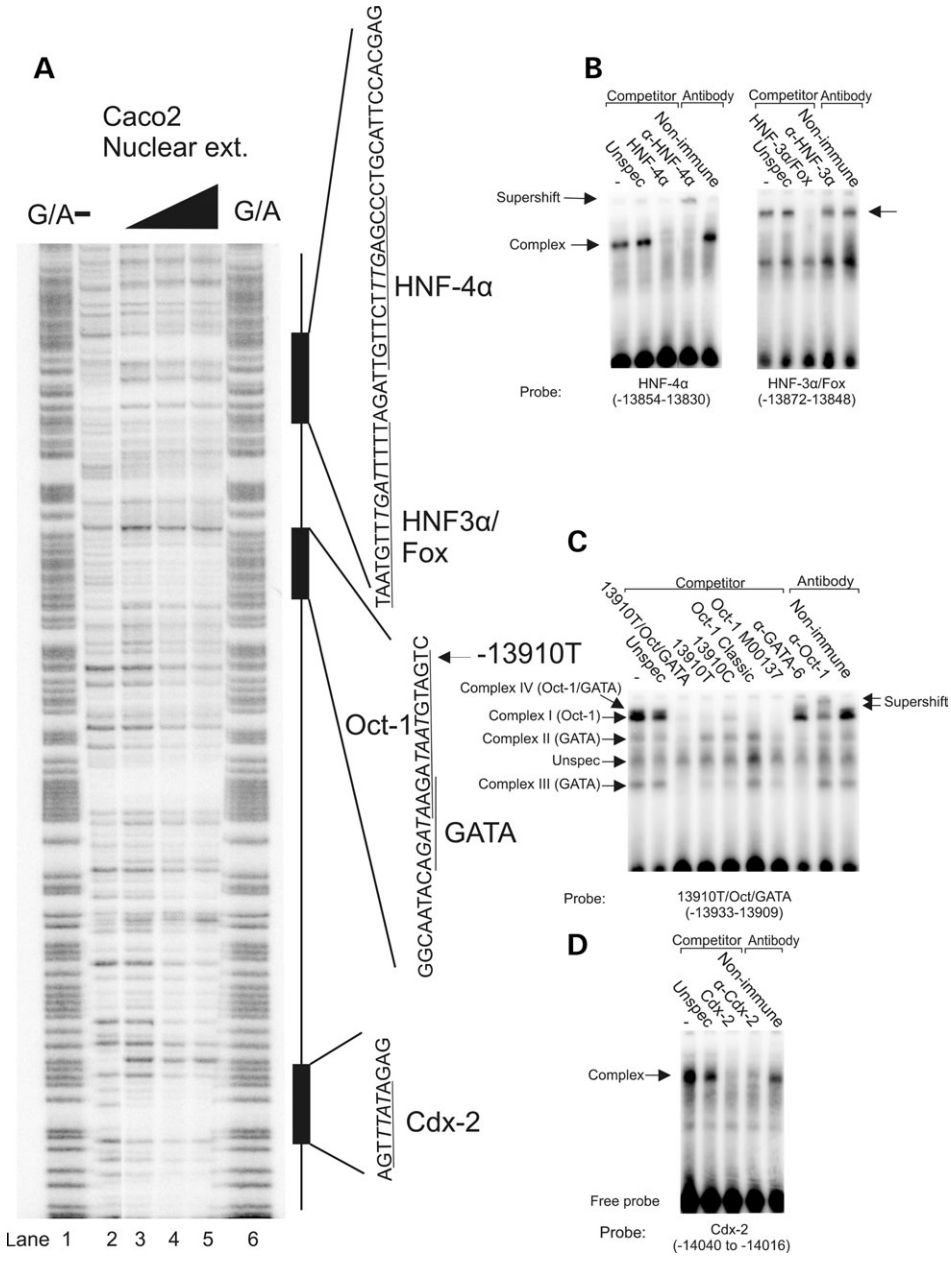

Figure 2. Analyses of protein/DNA interactions in the $T_{-13\,910}$ enhancer region. (**A**) DNase I footprint analysis of the $T_{-13\,910}$ enhancer region. Lanes 1 and 6, G/A sequencing lanes used as marker to correlate the footprints to the sequence. Lanes 2–5, increasing amount of nuclear extract from differentiated Caco-2 cells was added (lane 2, no extract; lane 3, 50 μg; lane 4, 100 μg and lane 5, 150 μg). The sequences of three protected regions (footprints) are indicated. The transcription factor binding sites are underlined. The sequences in italics were mutated and analysed by transfection experiments (Fig. 3B). (**B**–**D**) Gel shift assays analysing interactions of nuclear factors from nuclear extracts from differentiated Caco-2 cells to the regions that are protected in the footprint analysis (Fig. 2A).

$T_{-13\,910}$-enhancer activity (Fig. 3A). The Oct-1 mutation resulted in 5-fold reduction of the enhancer activity of the $T_{-13\,910}$ enhancer, demonstrating a functional importance of the Oct-1 site (Fig. 3B). Similar analyses of the GATA, HNF4α and Fox sites in the $T_{-13\,910}$ enhancer (Fig. 3A) also demonstrated that these sites are important for full enhancer activity (Fig. 3B). However, mutations in the Cdx-2 site did not significantly change the enhancer activity (Fig. 3B).

Expression plasmids for Oct-1, GATA-6, HNF4α, Cdx-2, HNF1α and HNF1β were co-transfected with the LPH-promoter construct (pGL3-hLPH1085), the LPH-promoter/$T_{-13\,910}$ enhancer construct (pGL-3 hLPH1085-14T) and the LPH-promoter/$C_{-13\,910}$ enhancer construct (pGL-3 hLPH1085-14C) (Fig. 3C). Binding sites for Oct-1 and HNF4α are only present in the −13 910 enhancers and an HNF1 binding site is only present in the LPH promoter, whereas GATA-6 and Cdx-2 binding sites are present both in promoter and in the −13 910 enhancer (Fig. 3A). GATA-6, Cdx-2 and HNF1α resulted in an increased proximal promoter activity, which is in agreement with the previous reports

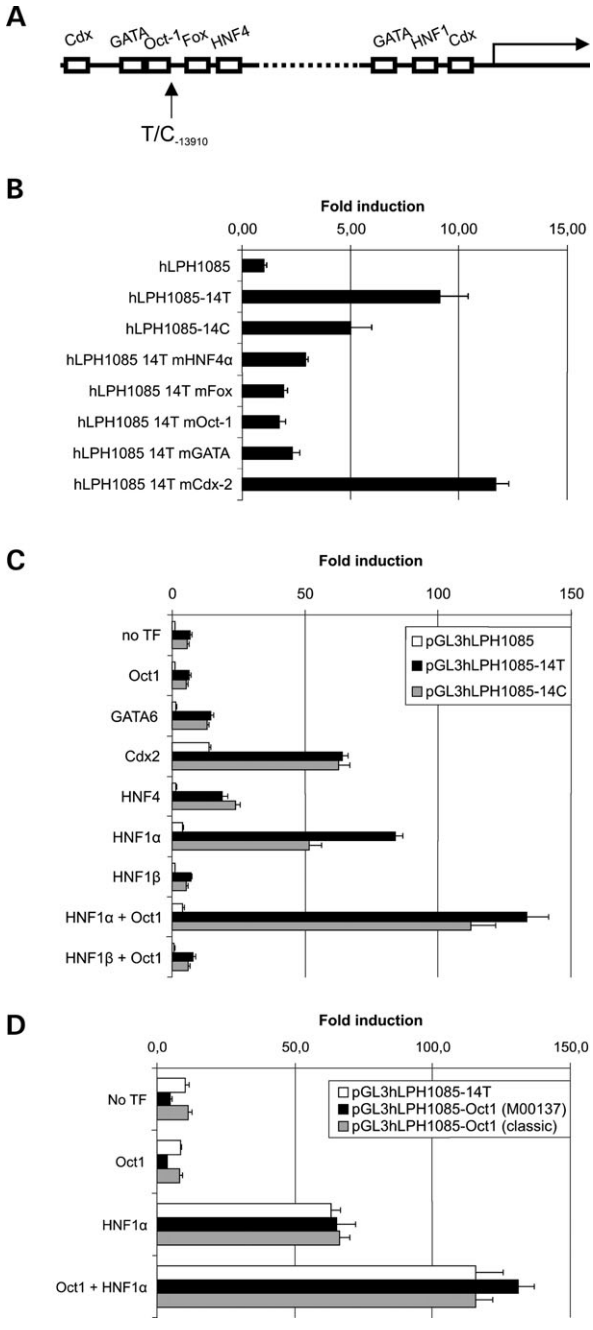

Figure 3. Functional analysis of the −13 910 enhancer region. (**A**) Structure of the −13 910 enhancer region and the LPH promoter showing the binding sites of HNF4α, Fox, Oct-1, GATA and Cdx-2 in the enhancer and the GATA, HNF1 and Cdx-2 sites in the proximal promoter. (**B**) Transfection analysis of the transcription factor binding sites in the −13 910 enhancer. The binding sites for HNF4α, Fox, GATA-6, Oct-1 and Cdx-2 were mutated in order to analyse the functional importance of these sequences. The positions mutated are marked in italics in Figure 2A. The luciferase activity was corrected for transfection efficiency and normalized to the basal expression of the human LPH promoter (pGL3-hLPH1085), $N = 4$. (**C**) Co-transfection of human LPH promoter constructs with expression plasmids for Oct-1, GATA-6, Cdx-2, HNF4α, HNF1α and HNF1β. 'No TF' indicates that no transcription factor expression plasmid was co-transfected. The luciferase activity was corrected for transfection efficiency and normalized to the expression of pGL3-hLPH1085, $N = 4$. (**D**) Analyses of the effect of changing the $T_{-13\ 910}$ Oct-1 site to the classical Oct-1 site and the M00137 Oct-1 site. The luciferase activity was corrected for transfection efficiency and normalized to the expression of pGL3-hLPH1085, $N = 4$.

(10,17,18,28). Oct-1 and HNF4α co-expression did not influence the proximal promoter activity, which is expected, as no binding site for these factors is present in the proximal promoter region. Surprisingly, co-expression of Oct-1 did not result in an increased $T_{-13\ 910}$ enhancer activity despite the presence of the Oct-1 binding site. HNF4α co-expression increased both the $T_{-13\ 910}$ and the $C_{-13\ 910}$ enhancer activities, whereas GATA-6 co-expression only slightly influenced the −13 910 enhancer activities. Cdx-2 strongly activates the proximal promoter activity (14-fold), and the presence of both −13 910 enhancers further increases the reporter gene expression to 63-fold. HNF1α has been shown to be important for LPH promoter activity and to interact with Cdx-2 (12), GATA factors (8–10) and Oct-1 (29). Activation of the LPH promoter activity by HNF1α is most likely mediated through binding to the proximal LPH promoter, as no HNF1 motif has been identified in the −13 910 enhancer region. It is therefore surprising that co-expression of HNF1α increases the reporter gene expression of pGL-3 hLPH1085-14T to 84-fold and that of pGL-3 hLPH1085-14C to 64-fold when compared with the activity of the basal LPH promoter. This finding suggests that the −13 910 enhancer effect is mediated through HNF1α bound at the proximal promoter. Indeed, co-expression of both HNF1α and Oct-1 further increases the effect of the $T_{-13\ 910}$ enhancer to 133-fold over the proximal promoter activity. This effect is specific to HNF1α, as Oct-1 does not synergize either with HNF1β, which also binds to the HNF1 motif in the LPH promoter (14) (Fig. 3C), or with Cdx-2, GATA-6 or HNF4α (data not shown). The Oct-1/HNF1α co-expression increases the $C_{-13\ 910}$ enhancer activity (112-fold). The lower activity of the $C_{-13\ 910}$ enhancer correlates with the weaker Oct-1 binding to the $C_{-13\ 910}$ SNP sequence (Fig. 1D) (6).

To further investigate the presence of a functional Oct-1 binding site in −13 910 enhancer, we changed the −13 910 Oct-1 site to the classical octamer Oct-1 site and to the non-classical M00137 Oct-1 site. Both these Oct-1 sites are able to sustain the high activity of the −13 910 enhancer, and the reporter gene expression of both constructs was only activated by Oct-1 over-expression when Oct-1 was co-expressed with HNF1α (Fig. 3D). These results validate the finding that the −13 910 Oct-1 site is a functional Oct-1 site.

The −13 910 SNP is located within intron 13 of the mini-chromosome maintenance protein 6 (MCM6) gene (4). To investigate whether the −13 910 enhancer is specific to LPH expression, the LPH promoter in pGL3 hLPH1085-13910T and pGL3 hLPH1085-13910C was replaced with a 512 bp human MCM6 promoter fragment resulting in the plasmids pGL3 hMCM6-13910T and pGL3 hMCM6-13910C, respectively. These were compared with pGL3-hMCM6 only containing the MCM6 promoter in transfection experiments using Caco-2 cells. The MCM6 promoter activity was not affected by either enhancer (data not shown), indicating that the −13 910 enhancer has a promoter preference and does not activate the MCM6 promoter.

DISCUSSION

We have identified Oct-1 as the transcription factor binding strongly to the $T_{-13\ 910}$ variant. The Oct-1 site is required

for full $T_{-13\,910}$ enhancer activity in combination with GATA-6, HNF4α and Fox sites. Mutation of either of these binding sites reduces the enhancer activity significantly. Over-expression of Oct-1 does not increase reporter gene expression of the −13 910 enhancer construct. A possible explanation could be that Oct-1 is highly expressed in Caco-2 cells and is not the limiting factor for the LPH expression. In order to detect an effect on the reporter gene expression of Oct-1, it is necessary to co-express it with HNF1α. HNF1α binding sites have not been identified in the −13 910 enhancer region, but the presence of the −13 910 enhancer increases the effect of the HNF1α over-expression (Fig. 3). Both $T_{-13\,910}$ and $C_{-13\,910}$ enhancer activities increase with HNF1α over-expression. The $T_{-13\,910}$ enhancer still induces a higher level of reporter gene expression than the $C_{-13\,910}$ enhancer when HNF1α is over-expressed, whereas over-expression of HNF4α, GATA-6 and Cdx-2 increases the enhancer but diminishes the difference between the $T_{-13\,910}$ and the $C_{-13\,910}$ enhancers, indicating that these factors are not directly involved in the differential activation by the $T_{-13\,910}$ and the $C_{-13\,910}$ enhancers. Over-expression of both Oct-1 and HNF1α further increases the reporter gene expression of the −13 910 enhancers, and we suggest that this synergistic effect of Oct-1/HNF1α expression is most likely mediated through interactions between Oct-1 bound at the $T/C_{-13\,910}$ in the enhancer and HNF1α bound at the well-characterized HNF1 site (10,12,14) in the proximal LPH promoter. Oct-1 and HNF1α have previously been reported to interact and affect intestinal and liver gene expression (29,30). Mutations in the GATA, HNF4, Fox or Oct-1 sites abolish the activity of the −13 910 enhancer. It is therefore clear that the −13 910 enhancer has a refined structure that requires the presence of the all four binding sites in order to possess enhancer activity. The $T_{-13\,910}$ sequence increases the binding of Oct-1 when compared with the $C_{-13\,910}$ sequence, and the Oct-1 binding is correlated to an increased enhancer activity of the $T_{-13\,910}$ variant.

In most mammals, the postweaning decline probably is an advantage as it forces the young mammal to be weaned from the mother. In humans, however, lactase persistence makes it possible to exploit milk from domestic animals. Lactase persistence is a late adaptation in the human evolution (3), and we suggest that a C to T substitution at position −13 910 has resulted in an increased activity of the −13 910 enhancer by creating a strong Oct-1 binding site. It is generally accepted that lactase persistence is a result of bypassing the postweaning decline of lactase that occurs in lactase non-persistent individuals and in other mammals. The mechanism behind the postweaning decline of lactase in mammals has not been identified, but it is known that it is a result of a decrease in the lactase promoter activity as relatively short LPH promoter fragments from pig (20) and rat (22) are able to direct a postweaning decline of a reporter gene in transgenic mice. We hypothesize that the postweaning decline could be either a result of a decrease in the recruitment of transcriptional activators or an increased recruitment of repressors to the LPH promoter. However, high activity of both pig and rat LPH promoters in the transgenic mouse models is dependent on enhancer regions (11,20–22,31). The enhancer regions are not conserved among pig, rat and human in contrast to the proximal promoter regions (1). As both the $T_{-13\,910}$ variant and the $C_{-13\,910}$ variant act as enhancers, it is possible that they both activate LPH expression after birth, but the presence of Oct-1 at the $T_{-13\,910}$ enhancer during the postweaning decline could result in an inability to down-regulate the lactase expression due to increased recruitment of transcriptional activators or by preventing the action of a repressor.

Interestingly, we have noted that not only lactase persistence but also γ-globin persistence is associated with a mutation in an Oct-1 site. γ-globin persistence can be caused by a C to T substitution in an Oct-1 motif in the human γ-globin promoter (32). This mutation abolishes a repression of the γ-globin gene, resulting in γ-globin expression after birth. Although different mechanism(s) seem to be involved in the development of lactase and γ-globin persistence, it is noteworthy that Oct-1 is involved in persistent expression both in the intestinal epithelium and in the haematopoietic system.

Oct-1 binding to the $T_{-13\,910}$ variant is necessary for the enhancer activity, and it is possible that GADPH is a coactivator for Oct-1, as it is the case for histone H2B promoter activity (26). Although we have not investigated the role of GAPDH for the enhancer activity, it is intriguing to speculate that the metabolic state of the cell might feed back and influence the transcription of the lactase gene. Oct-1 has been reported to recruit chromatin modifying co-factors, which are able to either enhance or silence the gene depending on the cell type and promoter architecture (33,34). Thus, it is possible that Oct-1 binding to the $T_{-13\,910}$ variant *in vivo* induces chromatin changes close to the *LCT* gene that are involved in the lactase-persistent phenotype.

MATERIALS AND METHODS

Fractionation of HeLa-S nuclear extract

We have previously shown that the factor binding $T_{-13\,910}$ is expressed in HeLa cells (6). HeLa S3 nuclear extracts were therefore used as starting material for the purification. HeLa S3 cells were grown in Joklik-modified minimal essential medium containing 5% fetal calf serum, 100 mg/ml of penicillin and streptomycin and 2 mmol/l L-glutamine in spinner bottles. Cells from 60 l of medium were harvested by centrifugation, and a nuclear extract was prepared (35). The nuclear extract was applied to a 75 ml Q-sepharose Fast Flow column (Amersham Biosciences) equilibrated with buffer D [20 M HEPES, pH 7.9; 20% glycerol; 1.5 mM $MgCl_2$; 100 mM KCl; 0.2 mM EDTA and 0.5 mM dithiothreitol (DTT)]. Bound proteins were eluted with buffer D supplemented with 1 M NaCl. The eluate was dialysed against buffer D, and NaCl was added to 50 mM and loaded to a Q-Sepharose column again. The flowthrough was collected and concentrated 10 times by ultrafiltration using a PM10 filter (Amicon). An Aliquot of 400 μl of the concentrated eluate was mixed with 800 μl buffer D and 1600 μl gel shift buffer [25 mM Tris, pH 7.8, 5 mM $MgCl_2$, 6 mM KCl, 0.5 mM EDTA, 1 mM DTT, 0.5 mM phenylmethylsulphonyl fluoride (PMSF), 5% Ficoll and 2.5% glycerol] and incubated for 10 min at 4°C. The sample was mixed with 1 mg of

streptavidin-coated magnetic beads (Dynal) coupled with a double-stranded biotinylated oligonucleotide containing three tandem-repeated sequences of a 24 bp region surrounding the $T_{-13\,910}$ position and incubated for 30 min at 4°C. The magnetic beads were collected using Magnetic Particle Concentrator (Dynal) and washed by resuspending the beads in 300 μl buffer D and 600 μl gel shift buffer. Bound proteins were eluted twice from the magnetic beads by resuspending the magnetic beads in buffer D containing 400 mM NaCl.

Samples taken during the different steps in the purification procedure were separated on a NuPAGE 12% Bis–Tris gel (Invitrogen). After electrophoresis, the gel was silver stained using the Bio-Rad Silver Stain Kit (Bio-Rad). Bands were excised and in-gel digested with trypsin (23). The proteolytic fragments were separated and sequenced by nanoflow liquid chromatography tandem mass spectrometry (LC MS/MS), using a Waters/Micromass QTOF Micro (24). The results from the mass spectrometry were utilized to search an in-house version of the Mascot search engine 2.0 (http://www.matrixscience.com/). The prominent ~35 kDa protein was identified as GAPDH (P04406) with a MOWSE score of 209 and 9 proteolytic fragment spectra assigned within 40 p.p.m. mass accuracy.

Fractionation of Caco-2 nuclear extract

The $T_{-13\,910}$-binding activity was purified from Caco-2 nuclear extract with a simpler procedure. About 1000 μg of nuclear extract isolated from differentiated Caco-2 cells (35) was mixed with 800 μl buffer D, 2000 μl gel shift buffer and 20 μg dI–dC, and the $T_{-13\,910}$-binding protein was affinity-purified as described with the Q-purified HeLa-S extract.

In total, 25 μl (~8 μg) of the diluted Caco-2 nuclear extract used in purification was separated on a NuPAGE 12% Bis–Tris gel (Invitrogen) together with samples from the steps of the fractionation of the Caco-2 nuclear extract. After electrophoresis, the gel was electrotransferred onto Immobilon membrane (Millipore). Immunoblotting was performed with primary antibodies to either Oct-1 (Santa Cruz) or GAPDH (Chemicon). The blot was developed using the ECL kit (Amersham Biosciences), and the chemiluminescence signals were captured using an LAS-1000+ (Fujifilm).

DNA footprint and gel shift assays

A fragment covering the $T_{-13\,910}$ variant labelled with ^{32}P at position −14 096 (at an internal *Eco*RI site) was used in DNase I footprint analysis. An aliquot of 10–30 μl of (50–150 μg) nuclear extract from differentiated Caco-2 was mixed with 20 μl gel shift buffer containing 1 mM DTT and Protease inhibitor cocktail (Sigma) (1 μl/ml) and 1 μg dI–dC (Amersham Biosciences) and was incubated for 10 min on ice. Five nanograms of ^{32}P-labelled $T_{-13\,910}$ fragment was added and incubated for 15 min on ice. About 500 ng of DNase I (Fermentas) was added and incubated for 60 s at room temperature. The DNase digestion was stopped by adding 300 μl stop buffer [1% sodium dodecyl sulphate (SDS), 0.33 M NaAc and 5 mM EDTA]. The samples were extracted with phenol/chloroform and ethanol precipitated. The precipitated DNA was analysed by denaturating polyacrylamide gel electrophoresis (PAGE) (6% and 6 M urea) followed by autoradiography.

Gel shift assays and supershifts were performed as described previously (21) using double-stranded oligonucleotides from the −13 910 enhancer region covering transcription factor binding sites: HNF4α (position −13 854 to −13 830; ttagattgttctttgagccctgcat), HNF3α/Fox (−13 872 to −13 848; ttgtataatgtttgatttttagatt), $T_{-13\,910}$/Oct/GATA (−13 933 to −13 909; gcaatacagataagataatgtagtc), 13910T (−13 922 to −13 898; aagataatgtagTccctggcctcaa), 13910C (−13 922 to −13 898; aagataatgtagCccctggcctcaa), Cdx-2 (−14 040 to −14 016; cacgtcatagtttatagagtgcata). Oct-1 classic (tgtcgaATGCAAATcactagaa) contained the canonical octamer sequence. Oct-1 M00137 motif (agctgaATATTAATCATAGtagctt) is one of the binding sites (R07117 from the Transfac database) that were used to construct the M00137 Oct-1 matrix. The Oct-1 site was flanked with random sequence to adjust the length of the oligonucleotides to the same length as the −13 910 enhancer oligonucleotides. All gel shift assay oligonucleotides were synthesized with a 5′ A-overhang to facilitate labelling by a forward reaction using polynucleotide kinase and [γ-P^{32}]ATP. For the supershift assays, the following antibodies were used: anti-HNF4α, anti-HNF3α, anti-Oct-1, anti-GATA-6 (Santa Cruz) and anti-Cdx-2 [a gift from Dr Michael German (36)].

Plasmid constructs and transfection

pGL3 SI257-13910T (6) was used as template to generate mutations in a 455 bp region around position −13 910 (−14 133 to −13 684). Mutations were introduced by polymerase chain reaction (PCR) mutagenesis by overlap extension (35). An *Xba*I site was introduced in the HNF4α motif by changing TCTTTG to TCTAGA at position −13 845 to −13 840. Likewise, *Xba*I sites were introduced into the Fox motif (TTTGAT → TCTAGA, −13 862 to −13 857), the GATA motif (AGATAA → TCTAGA, −13 926 to −13 921) and the Cdx motif (TTTATA → TCTAGA, −14 030 to −14 025). At the Oct-1 motif, an *Eco*RI site was introduced (GATAAT → GAATTC, −13 920 to −13 915), as an *Xba*I site would not destroy the Oct-1 consensus motif. All the mutations in the transcription binding sites significantly reduced the protein/DNA interactions to the sites when they were analysed in gel shift assays (data not shown). The $T_{-13\,910}$ Oct-1 site was furthermore changed to the classical canonical octamer sequence (ATGTAGT to TGCAAAT; −13 916 to −13 910) and to the non-classical M00137 Oct-1 (GATAATGTAGTCC → ATATTAATCATAG; −13 920 to −13 908). The mutated 455 bp PCR fragments were TA-cloned into pCR 2.1 (Invitrogen), and it was verified that fragments had no PCR-introduced mutations by sequencing. The mutated fragment cut out from the pCR 2.1 plasmid by a *Xho*I digestion and was cloned into the *Sal*I site in pGL3 hLPH1085 (6). The pGL3 hLPH1085 construct contains a 1085 bp fragment (position −1097 to −13) of the lactase promoter cloned into the *Sac*I/*Xho*I site in pGL3-basic. Caco-2 cells were grown, and transfections were performed as previously described (21). Expression plasmid for GATA-6 was kindly provided by Dr Steven Krasinski (37), Oct-1 by

Dr Winship Herr (38), HNF4α by Dr Frances Sladek (39) and Cdx-2 by Dr Michael German (36).

MCM6 promoter plasmids were constructed by PCR amplifying the −456 to +55 of the human MCM6 gene. This fragment was cloned into pGL3-basic (Promega) yielding the plasmid pGL3 hMCM6. Plasmids containing both the MCM6 promoter and the −13 910 enhancer were made by replacing the LPH promoter in pGL3 hLPH1085-13910T and pGL3 hLPH1085-13910C (6) with the MCM6 promoter fragment (pGL3 hMCM6 -13910T and pGL3 hMCM6 -13910C).

ACKNOWLEDGEMENTS

We thank Hans Sjöström for valuable comments and discussion of the manuscript. The work was support by the Lundbeck Foundation, the Novo Nordisk Foundation, the Foundation of 17.12.1981 and the Danish Medical Research Council.

Conflict of Interest statement. None declared.

REFERENCES

1. Troelsen, J.T. (2005) Adult-type hypolactasia and regulation of lactase expression. *Biochim. Biophys. Acta*, **1723**, 19–32.
2. Wang, Y., Harvey, C.B., Pratt, W.S., Sams, V.R., Sarner, M., Rossi, M., Auricchio, S. and Swallow, D.M. (1995) The lactase persistence/non-persistence polymorphism is controlled by a *cis*-acting element. *Hum. Mol. Genet.*, **4**, 657–662.
3. Bersaglieri, T., Sabeti, P.C., Patterson, N., Vanderploeg, T., Schaffner, S.F., Drake, J.A., Rhodes, M., Reich, D.E. and Hirschhorn, J.N. (2004) Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.*, **74**, 1111–1120.
4. Enattah, N.S., Sahi, T., Savilahti, E., Terwilliger, J.D., Peltonen, L. and Jarvela, I. (2002) Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.*, **30**, 233–237.
5. Mulcare, C.A., Weale, M.E., Jones, A.L., Connell, B., Zeitlyn, D., Tarekegn, A., Swallow, D.M., Bradman, N. and Thomas, M.G. (2004) The T allele of a single-nucleotide polymorphism 13.9 kb upstream of the lactase gene (*LCT*) (C-13.9kbT) does not predict or cause the lactase-persistence phenotype in Africans. *Am. J. Hum. Genet.*, **74**, 1102–1110.
6. Troelsen, J.T., Olsen, J., Møller, J. and Sjöström, H. (2003) An upstream polymorphism associated with lactase persistence has increased enhancer activity. *Gastroenterology*, **125**, 1686–1694.
7. Olds, L.C. and Sibley, E. (2003) Lactase persistence DNA variant enhances lactase promoter activity *in vitro*: functional role as a *cis* regulatory element. *Hum. Mol. Genet.*, **12**, 2333–2340.
8. Van Wering, H.M., Bosse, T., Musters, A., de Jong, E., de Jong, N., Hogen Esch, C.E., Boudreau, F., Swain, G.P., Dowling, L.N., Montgomery, R.K. *et al*. (2004) Complex regulation of the lactase-phlorizin hydrolase promoter by GATA-4. *Am. J. Physiol. Gastrointest. Liver Physiol.*, **287**, G899–G909.
9. Krasinski, S.D., Van Wering, H.M., Tannemaat, M.R. and Grand, R.J. (2001) Differential activation of intestinal gene promoters: functional interactions between GATA-5 and HNF-1 alpha. *Am. J. Physiol. Gastrointest. Liver Physiol.*, **281**, G69–G84.
10. Van Wering, H.M., Huibregtse, I.L., Van Der Zwan, S.M., De Bie, M.S., Dowling, L.N., Grand, R.J. and Krasinski, S.D. (2002) Physical interaction between GATA-5 and HNF-1alpha results in synergistic activation of the human lactase-phlorizin hydrolase promoter. *J. Biol. Chem.*, **277**, 27659–27667.
11. Verhave, M., Krasinski, S.D., Christian, S.I., Van Schaik, S., Van Den Brink, G.R., Doting, E.M., Maas, S.M., Wolthers, K.C., Grand, R.J. and Montgomery, R.K. (2004) Regulatory regions in the rat lactase-phlorizin hydrolase gene that control cell-specific expression. *J. Pediatr. Gastroenterol. Nutr.*, **39**, 275–285.
12. Mitchelmore, C., Troelsen, J.T., Spodsberg, N., Sjöström, H. and Norén, O. (2000) Interaction between the homeodomain proteins Cdx2 and HNF1alpha mediates expression of the lactase-phlorizin hydrolase gene. *Biochem. J.*, **346**, 529–535.
13. Mitchelmore, C., Troelsen, J.T., Sjöström, H. and Norén, O. (1998) The HOXC11 homeodomain protein interacts with the lactase-phlorizin hydrolase promoter and stimulates HNF1alpha-dependent transcription. *J. Biol. Chem.*, **273**, 13297–13306.
14. Spodsberg, N., Troelsen, J.T., Carlsson, P., Enerback, S., Sjöström, H. and Norén, O. (1999) Transcriptional regulation of pig lactase-phlorizin hydrolase: involvement of HNF-1 and FREACs. *Gastroenterology*, **116**, 842–854.
15. Troelsen, J.T., Olsen, J., Mitchelmore, C., Hansen, G.H., Sjöström, H. and Norén, O. (1994) Two intestinal specific nuclear factors binding to the lactase-phlorizin hydrolase and sucrase–isomaltase promoters are functionally related oligomeric molecules. *FEBS Lett.*, **342**, 297–301.
16. Troelsen, J.T., Olsen, J., Norén, O. and Sjöström, H. (1992) A novel intestinal *trans*-factor (NF-LPH1) interacts with the lactase-phlorizin hydrolase promoter and co-varies with the enzymatic activity. *J. Biol. Chem.*, **267**, 20407–20411.
17. Troelsen, J.T., Mitchelmore, C., Spodsberg, N., Jensen, A.M., Norén, O. and Sjöström, H. (1997) Regulation of lactase-phlorizin hydrolase gene expression by the caudal-related homeodomain protein Cdx-2. *Biochem. J.*, **322**, 833–838.
18. Fang, R., Olds, L.C., Santiago, N.A. and Sibley, E. (2001) GATA family transcription factors activate lactase gene promoter in intestinal Caco-2 cells. *Am. J. Physiol. Gastrointest. Liver Physiol.*, **280**, G58–G67.
19. Fang, R., Santiago, N.A., Olds, L.C. and Sibley, E. (2000) The homeodomain protein Cdx2 regulates lactase gene promoter activity during enterocyte differentiation. *Gastroenterology*, **118**, 115–127.
20. Troelsen, J.T., Mehlum, A., Olsen, J., Spodsberg, N., Hansen, G.H., Prydz, H. and Sjöström, H. (1994) 1 kb of the lactase-phlorizin hydrolase promoter directs post-weaning decline and small intestinal-specific expression in transgenic mice. *FEBS Lett.*, **342**, 291–296.
21. Troelsen, J.T., Mitchelmore, C. and Olsen, J. (2003) An enhancer activates the pig lactase phlorizin hydrolase promoter in intestinal cells. *Gene*, **305**, 101–111.
22. Lee, S.Y., Wang, Z., Lin, C.K., Contag, C.H., Olds, L.C., Cooper, A.D. and Sibley, E. (2002) Regulation of intestine-specific spatiotemporal expression by the rat lactase promoter. *J. Biol. Chem.*, **277**, 13099–13105.
23. Stensballe, A. and Jensen, O.N. (2001) Simplified sample preparation method for protein identification by matrix-assisted laser desorption/ionization mass spectrometry: in-gel digestion on the probe surface. *Proteomics*, **1**, 955–966.
24. Hjerrild, M., Stensballe, A., Jensen, O.N., Gammeltoft, S. and Rasmussen, T.E. (2004) Protein kinase A phosphorylates serine 267 in the homeodomain of engrailed-2 leading to decreased DNA binding. *FEBS Lett.*, **568**, 55–59.
25. Mazzola, J.L. and Sirover, M.A. (2003) Subcellular localization of human glyceraldehyde-3-phosphate dehydrogenase is independent of its glycolytic function. *Biochim. Biophys. Acta*, **1622**, 50–56.
26. Zheng, L., Roeder, R.G. and Luo, Y. (2003) S phase activation of the histone H2B promoter by OCA-S, a coactivator complex that contains GAPDH as a key component. *Cell*, **114**, 255–266.
27. Verrijzer, C.P., Alkema, M.J., van Weperen, W.W., Van Leeuwen, H.C., Strating, M.J. and van der Vliet, P.C. (1992) The DNA binding specificity of the bipartite POU domain and its subdomains. *EMBO J.*, **11**, 4993–5003.
28. Van Wering, H.M., Moyer, L., Grand, R.J. and Krasinski, S.D. (2002) Novel interaction at the Cdx-2 binding sites of the lactase-phlorizin hydrolase promoter. *Biochem. Biophys. Res. Commun.*, **299**, 587–593.
29. Ishii, Y., Hansen, A.J. and Mackenzie, P.I. (2000) Octamer transcription factor-1 enhances hepatic nuclear factor-1alpha-mediated activation of the human UDP glucuronosyltransferase 2B7 promoter. *Mol. Pharmacol.*, **57**, 940–947.
30. Zhou, D.X. and Yen, T.S. (1991) The ubiquitous transcription factor Oct-1 and the liver-specific factor HNF-1 are both required to activate transcription of a hepatitis B virus promoter. *Mol. Cell Biol.*, **11**, 1353–1359.
31. Krasinski, S.D., Upchurch, B.H., Irons, S.J., June, R.M., Mishra, K., Grand, R.J. and Verhave, M. (1997) Rat lactase-phlorizin hydrolase/human growth hormone transgene is expressed on small intestinal villi in transgenic mice. *Gastroenterology*, **113**, 844–855.

32. Liu, L.R., Du, Z.W., Zhao, H.L., Liu, X.L., Huang, X.D., Shen, J., Ju, L.M., Fang, F.D. and Zhang, J.W. (2005) T to C substitution at −175 or −173 of the gamma-globin promoter affects GATA-1 and Oct-1 binding *in vitro* differently but can independently reproduce the hereditary persistence of fetal hemoglobin phenotype in transgenic mice. *J. Biol. Chem.*, **280**, 7452–7459.
33. Zabel, M.D., Wheeler, W., Weis, J.J. and Weis, J.H. (2002) Yin Yang 1, Oct1, and NFAT-4 form repeating, cyclosporin-sensitive regulatory modules within the murine CD21 intronic control region. *J. Immunol.*, **168**, 3341–3350.
34. Remenyi, A., Scholer, H.R. and Wilmanns, M. (2004) Combinatorial control of gene expression. *Nat. Struct. Mol. Biol.*, **11**, 812–815.
35. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (2002) *Current Protocols in Molecular Biology*. John Wiley & Sons, NY.
36. German, M.S., Wang, J., Chadwick, R.B. and Rutter, W.J. (1992) Synergistic activation of the insulin gene by a LIM-homeo domain protein and a basic helix–loop–helix protein: building a functional insulin minienhancer complex. *Genes Dev.*, **6**, 2165–2176.
37. Morrisey, E.E., Ip, H.S., Lu, M.M. and Parmacek, M.S. (1996) GATA-6: a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. *Dev. Biol.*, **177**, 309–322.
38. Cleary, M.A., Stern, S., Tanaka, M. and Herr, W. (1993) Differential positive control by Oct-1 and Oct-2: activation of a transcriptionally silent motif through Oct-1 and VP16 corecruitment. *Genes Dev.*, **7**, 72–83.
39. Sladek, F.M., Zhong, W.M., Lai, E. and Darnell, J.E., Jr. (1990) Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. *Genes Dev.*, **4**, 2353–2365.