4 Discussion

The *cis*-acting regulatory elements that govern gene transcription (e.g. enhancers, locus control regions, and negative regulatory elements) can be located throughout the gene locus (White and White, 2001). However, in many cases the proximal promoter region contains elements that are sufficient to mediate high-level, tissue-specific gene transcription. These proximal regulatory elements often contain binding sites for tissue-restricted proteins that mediate transcriptional activation in eukaryotic cells. In this work, the proximal 5′-flanking region of the LI-cadherin was isolated and characterized. LI-cadherin is a transmembrane glycoprotein with a restricted expression pattern. In mouse, LI-cadherin is solely expressed in the intestine, whereas in rat it is also expressed in the liver, hence the name LI (liver and intestine)-cadherin. The expression of other cadherins is not always restricted to one tissue; the best studied classical cadherin is E-cadherin, which is concentrated in adherens junctions in all epithelial cells. In the intestine, LI-cadherin is co-expressed with E-cadherin, but in contrast to E-cadherin, LI-cadherin is distributed evenly over the lateral cell membrane. Because of its intestine-specific expression, the mouse LI-cadherin promoter is a useful tool for the study of transcriptional regulation.

4.1 Mapping of the transcription start site of the LI-cadherin gene

Before investigating the activity of the regulatory region, the promoter region has to be determined. Since the promoter is defined as the region directly upstream of the transcription start site, the transcriptional start site was mapped by analyzing the 5′-end of the LI-cadherin mRNA. Two methods were performed to determine the size of the mRNA. The primer extension method yielded an extension product which was only 7 nucleotides longer than the product received by 5′-RACE-PCR. Both results were consistent with the RNA size determination performed by René Jung (Jung, 2000).

Although the RACE-PCR technique has been successfully used to determine the transcription start sites in other genes (Nollet et al., 1996), the sequence information obtained from mRNA has one disadvantage: When the synthesis of the double strand is initiated, the first nucleotides at both ends are “turned down” so that a lariat of the mRNA is created. By preparation of blunt ends for the adapter ligation, the nucleotides in these lariats are lost.

Results from both the primer extension analysis and the 5′-RACE PCR were reproducible and consistent with each other, and further experiments to confirm the transcription start site were thus not performed.

The identified transcription initiation site resembled the consensus sequence C<u>A</u>G/TT that is described as an optimal initiation site for RNA polymerase II (Kraus et al., 1996). The underlined A is the first transcribed nucleotide. This consensus sequence is also supported by an analysis of 502 vertebrate mRNAs (Bucher, 1990), in which the consensus sequence TC<u>A</u>GT was found to resemble a common motif containing the transcription start site. Within this sequence, C<u>A</u> was the most frequently found nucleotide pair. The experimentally defined transcription start site A**C<u>A</u>G**G in the murine LI-cadherin gene matched the optimal motif in three nucleotides (in bold), and was therefore likely to be the transcriptional starting sequence. According to the results from the primer extension analysis, a second transcription start site

could possibly be located 18 bp further downstream within the sequence motif TCT. Since this region had no homology with the consensus sequence, the underlined adenine within the sequence AC<u>A</u>GG was defined as the first transcribed nucleotide (+1), and the 5′-flanking region was defined as the murine LI-cadherin promoter.

4.2 Verification and examination of LI-cadherin promoter activity

Reporter gene assays were performed in order to verify the transcriptional activity of the 5′-flanking region of the LI-cadherin gene. 1.8 kb and 2.8 kb of the murine 5′-flanking region, including the transcription initiation site and 67 bp of the untranslated exon 1, were ligated into a plasmid containing the enhanced green fluorescent protein (EGFP) as reporter gene. The reporter gene constructs were transfected into the mouse intestinal endocrinal cell line STC-1 that expresses endogenous LI-cadherin as shown by Western blotting.

Both the 2.8 kb and the 1.8 kb fragments of the LI-cadherin 5′-flanking region were able to drive GFP expression in STC-1 cells, indicating that the 1.8 kb fragment contains all regulatory elements sufficient for transcriptional activation. There was no detectable difference between the two LI-cadherin promoter constructs in the number of fluorescent cells and the apparent fluorescence intensity of single cells. Transfection with the positive control plasmid pEGFP-C1 containing a CMV promoter resulted in an increased number of cells expressing EGFP exhibiting also an increased fluorescence intensity. There was no detectable GFP-expression in cells transfected with the promoter-less plasmid pEGFP-1.

The GFP reporter gene system allows a qualitative control of transcription activity and a rough estimation of transfection efficiency. However, it was not possible to detect any differences in promoter activation between the 1.8 kb and 2.8 kb LI-cadherin constructs. For a quantitative analysis, it was necessary to use a system in which differences of transcriptional activation can be exactly determined. The 1.8 kb and 2.8 kb fragments containing the LI-cadherin 5′-flanking region were thus cloned into the plasmid pGL3-basic containing the coding region for the reporter gene firefly luciferase. Luciferase activity was normalized to transfection efficiency by using the dual-luciferase reporter assay system. First, the isolated regions were found to contain a functional activating region that was orientation-specific. Luciferase activity was 30% higher in cells transfected with pLI(1.8F)-luc compared to pLI(2.8F)-luc transfected cells. Each fragment was able to activate the transcription 33- and 23-fold compared to the 2.8 kb fragment cloned in the antisense orientation which served as a negative transcription control. Remarkably, transfection with the positive control plasmid pGL3-control containing the viral SV40 promoter resulted in only a 12-fold activation compared to the negative control. The linear relationship between firefly and *Renilla* luciferase activities indicates that the dual-luciferase reporter assay system is suitable for standardizing firefly luciferase activity to inherent assay-to-assay variances.

4.3 Sequence analysis of the 5′-flanking region of the LI-cadherin gene

A comparison of about 1500 nucleotides of the murine 5′-flanking region with the cloned 5′-flanking region of the human LI-cadherin gene (Gimpel, 2001), revealed that several regions contained sites with more than six homologous nucleotides. These sites may be important for the regulation of the restricted expression pattern of LI-cadherin in both species.

The conserved regions detected in both sequences were further examined for a potential regulatory role. Seventeen homologous regions were detected that are recognition sites for one or more transcription factors. The region from position –1505 to –811 contained nine conserved recognition sequences, followed by a region of about 540 nucleotides containing no homologous binding elements (figure 4.1). In the next 270 nucleotides, nine conserved sequences containing eighteen transcription factor-binding elements were detected. These observations suggest that the region from –1505 to –811 in the 5′-flanking region and -270 to +1 may be of importance for the transcriptional regulation of the LI-cadherin gene. In several genes, a TATA box located about 25 nucleotides upstream from the transcription start site is the most important sequence that signals the start of transcription (White and White, 2001). The LI-cadherin promoter is TATA-less but contains other typical eukaryotic promoter elements including a GC-box (–33 to –28) (Briggs et al., 1986), a CCAAT binding site (–86 to –82) (Bucher, 1990), and a GAGA-element (–16 to –26) (Legraverend et al., 1996; Wyse et al., 2000). In the human sequence, the GC-box is missing. The GC-box is the binding site for Sp-factors, which are ubiquitously expressed zinc finger proteins that can initiate transcription by recruiting the RNA polymerase holoenzyme to the promoter (Bucher, 1990). In addition, GC-boxes represent binding sites for tissue-specific Krüppel-like factors (Shields and Yang, 1998) that belong to a family of zinc finger proteins that are homologous to the *Drosophila melanogaster* segmentation gene product Krüppel (Schuh et al., 1986).

A purine-rich region containing alternating guanidines and adenines is conserved directly upstream of the transcription start site in the LI-cadherin 5′-flanking region. This so-called GAGA-motif is a novel *cis*-acting element that mediates basal and growth factor-dependent gene transcription (Legraverend et al., 1996; Wyse et al., 2000). Functional GAGA-motifs have been identified in the promoter regions in a number of genes in different species including the *Drosophila* Krüppel (Kerrigan et al., 1991) and heat shock proteins (Lee et al., 1992). These GAGA-motifs possess GA repeats; however, the overall sequence differs between the genes and species. Numerous studies have demonstrated that a GAGA-motif can be involved in transcription initiation with or without a functional TATA-element. For example, in the hsp71 promoter of *Drosophila melanogaster*, the GAGA-motif is required for recruitment of the TATA box binding protein (TBP) to its TATA-motif (Weber et al., 1997). Alternatively, a number of GAGA-motifs found in TATA-less promoters of other genes are necessary for the initiation of transcription (Bossone et al., 1992).

A modified CCAAT-element was detected in the region from –86 to –82. The CCAAT-motif is a common promoter element present in the proximal promoter of numerous mammalian genes (Bucher, 1990), and it is most often found between 80 and 100 bp upstream of the transcription start site. Several proteins have been isolated that bind to CCAAT-motifs, including CBP (also called NF-Y and CP1) which requires a highly conserved CCAAT-sequence. Other CCAAT-binding proteins include CTF/NF1 and the CCAAT/enhancer-binding protein (C/EBP), but DNA sequences in the binding sites of these proteins often do not contain a complete CCAAT-motif. In many cases, the CCAAT-motif co-operates with other specific elements to stimulate transcription (Jackson et al., 1995).

The presence of binding sites for homeodomain transcription factors - the caudal-related homeodomain protein Cdx2 and hepatocyte nuclear factor-1 (HNF-1) - is a shared feature of

several intestine specific expressed genes. These include lactase-phlorizin hydrolase (LPH) and sucrase isomaltase (SI) (Spodsberg et al., 1999; Wu et al., 1994).

A binding site for HNF-1 in position –52 to –40, as well as a binding site for Cdx2 in position –97 to –91 are conserved in the LI-cadherin 5′-flanking region. In addition, a second conserved Cdx2 binding site was detected within the region from –66 to –61 on the antisense strand.

Cdx2 and HNF-1α are both expressed in epithelial cells in the small intestine from an early stage in development, This shows their important role in intestinal development (Beck et al., 1995). The importance of Cdx2 for intestinal cell differentiation has been demonstrated by the conditional expression of Cdx2 in undifferentiated IEC-6 intestinal cells, which leads to a differentiated phenotype (Suh and Traber, 1996). HNF-1α and HNF-1β are related transcription factors that bind to DNA as homo- or heterodimers (Mendel et al., 1991). They were first identified as liver-enriched transcription factors involved in the expression of several plasma proteins, including α_1-antitrypsin and clotting factors (Mendel et al., 1991).

Studies have demonstrated that the intestine-specific expression pattern of LPH is due to the simultaneous presence of Cdx2 and HNF-1α (Mitchelmore et al., 2000), suggesting that a synergy between Cdx2 and HNF-1α is important for gene expression in the intestine. In addition, HNF-1α, Cdx2, and Cdx1 as well as GATA-4 are the primary transcription factors from the adult mouse intestinal epithelium to interact with the SI promoter (Boudreau et al., 2002). GATA-4 requires the presence of both HNF-1α and Cdx2 to influence the SI promoter activity, implying a combinatory role of these factors for the regulation of SI transcription during intestinal development.

Four conserved GATA-factor binding sites were identified in the LI-cadherin promoter within the 5′-flanking region from –267 to –87. The GATA family of transcription factors has been implicated in cell lineage differentiation during vertebrate development. Defined by two evolutionarily conserved zinc fingers that mediate binding to the consensus DNA sequence WGATAR (where W = A or T, and R = A or G), the GATA family is generally categorized into two classes based on their expression patterns and amino acid homology (Ko and Engel, 1993; Merika and Orkin, 1993). The GATA-1, -2, and -3 subfamily proteins control critical steps in erythroid and lymphoid development (Fitzgerald et al., 1998; Orkin et al., 1992), whereas GATA-4, -5, and -6 have been shown to modulate promoter function of intestinal genes, including the rat and human LPH (Fang et al., 2001; Fitzgerald et al., 1998; Krasinski et al., 2001), human sucrase isomaltase (Krasinski et al., 2001), and Xenopus intestinal fatty acid-binding protein (Gao et al., 1998) genes.

Other conserved DNA-binding elements in the LI-cadherin 5′-flanking region included Oct-1, AP-1, RFX-1, cMyb, STAT1, IK2, and GFI1. Oct-1 is known as a nuclear protein ubiquitously expressed in all eukaryotic cells. It plays multiple roles, acting as either a positive or a negative regulator of gene transcription (Ryan and Rosenfeld, 1997; Verrijzer and Van der Vliet, 1993). AP-1 is a homo- or heterodimeric DNA-binding protein composed of either two Jun family proteins or one Jun and one Fos family protein via the interaction of the leucine zipper domains (Angel et al., 1991). AP-1 may be involved in induction of a variety of genes regulating cell proliferation and differentiation during embryogenesis as well

as tumorigenesis (Tufan et al., 2002). RFX-1 is a member of the RFX family (Reith et al., 1989) and is also a ubiquitously expressed protein (Reith et al., 1994). RFX-1 has been shown to increase transcription from hepatitis B virus enhancers and to play a role in the induction of MHC class II genes (Siegrist et al., 1993). C-myb belongs to a family of transcription factors that all recognize the same DNA sequence (Ganter et al., 1999). Evidence is accumulating that all are involved in the regulation of either proliferation or differentiation or both in different cell types, c-myb being more specific for hematopoietic cells (Golay et al., 1996). Stat1 is member of the ligand-activated transcription factor family that serves in the dual function of signal transducer and activator of transcription (Darnell et al., 1994). IK2 is an essential regulator in the development and homeostasis of the lymphopoietic system (Molnar et al., 1996). The transcriptional repressor GFI1 is a zinc finger protein expressed in T-cell precursors in the thymus and in activated mature T-lymphocytes (Karsunky et al., 2002). One or more of these factors are candidates for interacting with the LI-cadherin promoter and for regulating the restricted expression of LI-cadherin. The identification of putative *cis*-acting elements was refined further by analyzing the activities of a serial of deletion constructs.

Figure 4.1: Comparison of 1500 bp of the murine (upper sequence) and human 5′-flanking (lower sequence) of the LI-cadherin gene. The nucleotides differing between the species are shaded in gray. Lines mark potential binding motifs. +1 indicates the transcriptional start point of the murine gene.

4.4 Deletion analysis of the 5′-flanking region of the murine LI-cadherin gene

Deletion analysis was performed to identify regions of the LI-cadherin promoter required for activity in transfected intestinal cells. The results indicate that promoter activity was dependent on positive regulatory elements that were located within only 55 bp directly

upstream of the transcriptional start site. Deletion of 22 bp from the –55 bp construct (yielding a –33 bp construct) included the removal of the GC-box and the binding sites for HNF-1 and AP-1 genes, indicating a crucial role for these binding factors. However, only the murine LI-cadherin gene, but not the human one, contains a GC-box upstream of the transcription start site, suggesting that this binding site is not involved in the regulation of the specific expression of LI-cadherin. The perfect conservation of the HNF-1 and AP-1 binding elements in the LI-cadherin genes emphasizes the importance of HNF-1 and/or AP-1. Removal of the GAGA-element (-22 bp construct) did not further decrease the promoter activity. Thus, this element is not able to activate the transcription by itself.

Other regions also influenced the promoter activity negatively or positively. The removal of ten nucleotides from the –91 bp construct including the CCAAT-element resulted in a 45% decrease of luciferase activity, indicating that a CCAAT binding protein may be important for the positive regulation. Additional regions that positively regulated the transcription were detected in the positions –1200 to –1128 and –617 to –528, but these regions lack potential transcription factor binding sites and were not further analyzed. Interestingly, removal of thirteen nucleotides from the –81 bp construct resulted in a 100% increase of luciferase activity. An AP-1 binding site was completely deleted in the resulting –68 bp construct, indicating that a factor binding to this element may repress the transcriptional initiation.

Further studies in transgenic mice could be useful to verify whether the 55-bp region is sufficient to direct expression *in vivo* and to examine whether elements within this region are responsible for the tissue-specific expression of LI-cadherin. In order to identify proteins interacting with the first 100 nucleotides proximal to the transcription start site, EMSA was performed with nuclear extract isolated from the intestinal STC-1 cells.

4.5 DNA-protein interactions and influence of mutations on the promoter activity

Studies using EMSA demonstrated that nuclear extracts from STC-1 cells and murine enterocytes contained proteins that bind specifically to the functionally important region of the LI-cadherin promoter. These proteins are candidate transcription factors regulating LI-cadherin gene expression in intestinal cells. However, many transcription factors are expressed in several tissues, so that an identification of a tissue-restricted transcription factor would require an examination of a wide range of cell lines from different tissues.

Three regions of the murine promoter, including the nucleotides –100 to –71, –70 to –41, and –50 to –21 were examined for their ability to interact with nuclear proteins from cultured STC-1 cells as well as from murine enterocytes.

The mobility of the retarded bands with the oligonucleotide encompassing the region –50 to –21 was similar when using nuclear extracts from STC-1 cells or enterocytes except for two additional complexes detected only with enterocyte extracts. Specifically, a Sp1 factor seems to participate in the recognition of the GC-rich region within position –50 to –21 in the murine promoter, as was established by supershift analysis and mutational analysis. The same complex formation was also found with the GC-rich region present in the proximal 5′-flanking region of the E-cadherin gene, confirming that Sp1 can bind to both promoters. Involvement of Sp1 factors in the recognition of the GC-rich region of the E-cadherin and the

P-cadherin promoter has already been established; also a positive regulatory role for these elements in both promoters has been confirmed (Faraldo et al., 1997).

Two different mutations (mGC1 and mGC2) within the GC-box of the reporter gene construct pLI(1.8F)-luc and pLI(55)-luc resulted in a decrease of the promoter activity in STC-1 cells. However, while the mutation mGC1 in the longest promoter construct resulted in a complete loss of promoter activity, the activity of the 91 bp promoter containing the same mutation decreased only by 25%. An explanation for this can be that other factors binding upstream of nucleotide –91 are involved in repression of the transcription. While these binding sites are missing in the 91 bp promoter, their repressor-effect is lost. This could also mean that the basic promoter activity does not only depend on binding of Sp1. A co-operation between Sp1 and one or more other factors binding to the proximal 5′-flanking region could thus be responsible for the basic promoter activity. The significance of the GC-box should be carefully evaluated since the GC-box is lacking in the human sequence.

The –50 to –21 region also contains a putative binding site for the activator protein-1 (AP-1a). Supershift assay with an antibody raised against AP-1 (c-jun) did not change the retardation pattern, suggesting that another factor than AP-1 is eventually involved in complex formation.

Binding sites for HNF-1 and Cdx2 (Cdx2-a) are present within the region from –70 to –41. Only HNF-1 binds to this region, as demonstrated by supershift analysis and competition with oligonucleotides containing mutated HNF- and Cdx2-a binding sites. After introduction of these mutations into a reporter gene construct, the luciferase activity was completely lost compared with the wild-type promoter construct. This demonstrated that HNF-1 plays a crucial role in the regulation of the transcription of the LI-cadherin gene.

A second Cdx2 binding site (Cdx2-b) was found within the –100 to –71 region. Addition of anti-Cdx2 antibody resulted in a supershifted band demonstrating that Cdx2 binds to this site. Competition with an oligonucleotide containing only the conserved binding site for Cdx2 indicated that the lower part of complex 1 was made of Cdx2. Mutation of the Cdx2-b binding site resulted in loss of the upper part of complex 1 whereas the lower part remained. Addition of anti-AP-1 antibody and competition with mutated AP-1b and CAAT binding sites indicated that none of these sites are involved in protein interaction.

Mutation of the Cdx2-b binding site in the 1.8 kb promoter reporter gene construct resulted in a 40% decreased luciferase activity compared to the wild type promoter. Thus, these findings show that the distal Cdx2 binding site (Cdx2-b) and the HNF-1 binding site interact with proteins and positively regulate the transcription of the LI-cadherin gene.

Other regions of interest included the conserved nucleotides from –275 to –247 and the GAGA-element directly upstream of the transcription start site. Both were highly conserved in mouse and human. Deletion of the conserved sequence of the region from –275 to –247 containing two binding sites for GATA factors, one for cMyb, and one for Oct1, did not affect the promoter activity (not shown). Mutation of a sequence confirming the GAGA-motif located at position –30 resulted in a 40% decrease of promoter activity in Caco2 cells (not shown). However, by deletion analysis, it was shown that 33 bp directly upstream of the

transcription start site, including the GAGA-motif, were not able to drive the transcription alone.

Taken together, these data implicate the HNF-1 and the distal Cdx2 binding sites in regulating the transcriptional activity of the LI-cadherin gene in STC-1 cells. Figure 4.2 depicts schematically the first 270 bp of the murine and human transcription factor consensus sequences and reveals that the active Cdx2 and HNF-1 sites are conserved at identical positions in both sequences. This adds weight to the importance of HNF-1 and Cdx2 in regulating the LI-cadherin expression.

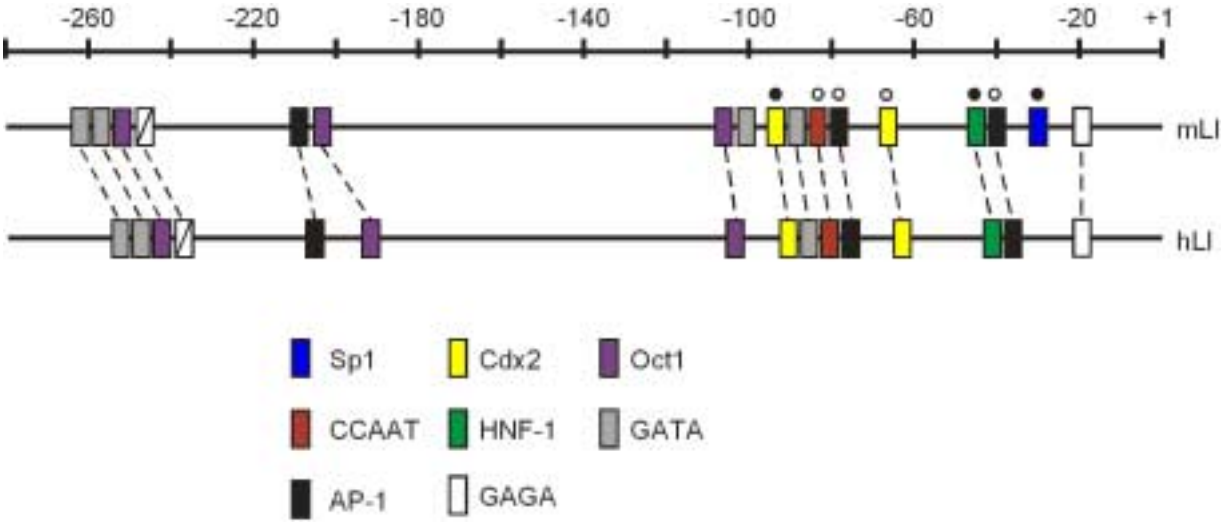

Figure 4.2: Comparison of the first 270 bp of the murine and human LI-cadherin promoter. The colored boxes indicate the positions of different transcription factor binding sites. Homologous binding sites in the human and murine sequences are indicated by a dotted line. The circles denote the binding analyzed in this work (black circles indicate positive binding, white circles negative ones).

The clustering of critical regulatory elements within 100 bp 5′ of the transcription start site in the LI-cadherin gene suggests that this region comprises the core promoter of the TATA-less LI-cadherin gene. Inspection of Cdx2 and HNF-1 binding sites in many other known target genes expressed by intestinal epithelium reveals very similar findings. Such genes include sucrase isomaltase (SI) (Traber et al., 1992; Wu et al., 1992), lactase-plorizin hydrolase (*LPH*) (Fang et al., 2000; Mitchelmore et al., 2000; Troelsen et al., 1997), and claudin-2 (Sakaguchi et al., 2002). In all cases, the promoters contain Cdx and HNF-1 binding sites within a region of 25-110 bp upstream of the transcription start site. As reported by Sakaguchi (Sakaguchi et al., 2002), HNF-1α is able to enhance claudin-1-promoter activity only in the presence of Cdx2. In addition, Mitchelmore (Mitchelmore et al., 2000) demonstrates that HNF-1α physically interacts with Cdx2 as well as with GATA family members (van Wering et al., 2002) to co-operatively activate the pig LPH-promoter. HNF-1α, GATA-4, and Cdx2 also

interact and co-operate for the transcriptional activation of the SI gene (Boudreau et al., 2002). Interestingly, HNF-1 family members have recently been demonstrated to direct kidney-specific expression of the Ksp-cadherin, which lacks a TATA box (Bai et al., 2002).

This study, along with those of others, reinforce the notion that Cdx2 and HNF-1 transcription factors play a pivotal role in determining the expression patterns of many intestine-specific genes. The role of Sp1 in the regulation of LI-cadherin expression is unclear since its binding site is completely lacking in the human LI-cadherin gene. Within the cadherin family, a GC-rich region and a CCAAT-box are identified as positive regulatory elements in the 5′-flanking region of the E- and P-cadherin gene (Faraldo et al., 1997; Hennig et al., 1996). The LI-cadherin promoter also contains a CCAAT-box, and the deletion analysis indicates that this element also acts as a transcriptional activator, although binding of protein to this site was not confirmed by EMSA. It is possible that a common mechanism for directing the RNA polymerase to the transcription start site is acting in the different cadherins, although these DNA-binding elements are not restricted to cadherin promoters. The different expression patterns of the different cadherins imply that one or more tissue-specific transcription factors, either alone or in co-operation, is responsible for the restricted expression of the different cadherins.

Generally, *in vitro* studies for the examination of regulatory mechanisms of promoters are limited in several respects. Binding sites for transcription factors might be detected that are unavailable in native chromatin, and binding sites for proteins that are present at low concentrations in nuclear extracts might be overlooked. In contrast, footprinting would reveal binding sites for transcription factors under physiological conditions. DNase I hypersensitive sites reflect chromatin regions which are transcriptionally active and free of typical nucleosomes, allowing transcription factors to bind (Elgin et al., 1988; Gross and Garrard, 1988). Furthermore, methylation of CpG sites in promoters has been shown to be an indication of gene silencing (Eden and Cedar, 1994; Tate and Bird, 1993). For example, the positive regulatory elements of the E-cadherin promoter (a CCAAT-box, E-Pal, and others) are specifically protected by transcription factors *in vivo* in well differentiated but not in poorly differentiated carcinoma cells (Hennig et al., 1995).

4.6 Tissue- and species-dependent activity of the LI-cadherin promoter

In order to analyze whether the activity of the LI-cadherin promoter is specific for intestine epithelial cells specific *in vitro*, different cell-lines were transfected with a reporter gene plasmid containing the murine LI-cadherin promoter. Fourteen cell lines derived from eight different tissues and three species were examined. The promoter was highly active in some selected cell lines originating from intestine, liver, pancreas, stomach as well as lung, whereas other cells from these tissues were not able to activate the promoter. An active promoter coincided with endogenous expression of LI-cadherin (shown by Western blot) in two of three intestinal cell lines (STC-1 and Caco2), whereas two other cell lines (IEC-6 and CMT93) neither activated the promoter, nor did they express endogenous LI-cadherin. A weak promoter activity was shown in HT29 cells negative for LI-cadherin.

The promoter was highly active in one of two human liver cell lines. Interestingly, the promoter activity was not consistent with expression of LI-cadherin; the luciferase positive

cell line (HepG2) was negative for LI-cadherin whereas the positive cell line (HuH7) expressed endogenous LI-cadherin. The gastric adenocarcinoma cell lines 23132/87 and AGS both expressed LI-cadherin, but the LI-cadherin promoter was active only in the 23132/87 cells. The LI-cadherin positive cells derived from pancreas (BON) and lung (A549) also activated the promoter. Three cell lines derived from kidney (293), hamster ovary (CHO), and skin (B16F0) were all LI-cadherin negative and did not activate the promoter.

Since 1.8 kb of the LI-cadherin promoter induce transcriptional activation in cells not containing endogenous LI-cadherin, this region may comprise a core promoter that contains transcriptional elements sensitive to activation by factors in LI-cadherin negative cells. In addition, silencer binding upstream of the investigated regulatory region may be necessary to direct tissue specific expression of LI-cadherin. Cells containing endogenous LI-cadherin which do not activate the promoter may lack transcriptional activating factors such as Cdx2 or HNF-1. Almost all of the tested cell lines are originating from tumors; they differ from normal cells and may not be representative for their original tissues. Reduced expression of E-cadherin, for example, is regarded as one of the main molecular events involved in dysfunction of the cell-cell adhesion system, triggering cancer invasion and metastasis (Birchmeier and Behrens, 1994). In human cancers, partial or complete loss of E-cadherin expression correlates with malignancy. Expression of LI-cadherin in human gastric carcinomas is another example of changed expression pattern of proteins in malignant cells (Grotzinger et al., 2001). Recent studies have shown that Cdx2 may play a role during intestinal metaplasia formation and gastric carcinogenesis. The observation that Cdx2 regulates the activity of the LI-cadherin promoter *in vitro* raises the possibility that dysregulated expression of LI-cadherin may also contribute to gastrointestinal tumors (Bai et al., 2002).

Thus, the expression of LI-cadherin in cells derived from tissues normally not containing endogenous LI-cadherin may be explained by their malignant origins. A prediction of a specific promoter activity was not possible after transfection of these cells. *In vivo* studies are necessary for testing tissue-restricted expression of LI-cadherin.

Notably, a species-dependent activity of the LI-cadherin promoter was detected *in vitro* with the human cell lines Caco2 and HepG2. The human promoter was 40- and 8-fold more active, respectively, than the murine promoter. In the murine cell line STC-1, the human promoter was 2.5-fold more active compared to the murine promoter, suggesting that the human 5′-flanking region generally has a higher capacity to activate transcription *in vitro*. The different regulatory strengths of 1.8 kb of the human and murine LI-cadherin 5′-flanking region might be due to the dissimilarities within this region. A clear difference is the lack of the GC-box in the human region. However, Sp1 binds to the murine GC-box as shown by EMSA, and mutations of this site down-regulated the promoter activity in reporter gene assays. Thus, these data suggest a positive regulatory role of the GC-box.

It is important to note that no individual factor is capable of playing a dominant role in generating the immense specificity required to regulate transcription in eukaryotes. Distinct multiprotein complexes are needed to modulate higher-order chromatin structure, to bind to promoters and enhancers, to communicate between activators and repressors and sites of

transcription initiation, to modify nucleosomal structure, and to generate transcripts. Each of these complexes might be a key player in regulating a given gene.

The finding that HNF-1, Cdx2, and Sp1 are able to activate the LI-cadherin promoter *in vitro* shows the importance of these factors for regulating the LI-cadherin gene. At the same time, it becomes obvious that the performed experiments can not fully describe the LI-cadherin promoter and that important factors may have been overlooked. In order to get further insight into the function of the identified factors for the tissue-specific expression of LI-cadherin, transgenic mice containing parts of the LI-cadherin promoter could be a powerful tool to analyze the role of these transcription factors *in vivo* and to determine the functional important regulatory region of the 5′-flanking region.