1 Introduction

1.1 Cell-cell adhesion

Cell-cell interactions are of fundamental importance for the development and the maintenance of tissues and organs in multicellular organisms. Cells adhere to each other and to the extracellular matrix through cell-surface proteins called cell adhesion molecules (CAMs). CAMs can be cell-cell adhesion molecules or cell-matrix adhesion molecules. Some CAMs are Ca^{2+}-dependent; these seem to be primarily responsible for the tissue-specific cell-cell adhesion seen in early vertebrate embryos (Gumbiner, 1996). Intercellular adhesion molecules are classified into four major molecular families: The selectin superfamily, the immunoglobulin superfamily, the integrin superfamily, and the cadherin superfamily.

Selectins are cell-surface carbohydrate-binding proteins that mediate a variety of transient, Ca^{2+}-dependent cell-cell adhesion interactions in the bloodstream (Osborn, 1990). The three other major classes of cell adhesion molecules function by protein-protein interactions. The majority of the known members of the integrin superfamily bind to various molecules of the extracellular matrix and mediate cell matrix binding. However, a few integrins may act as cell-cell adhesion molecules in a heterophilic fashion (Hynes and Gumbiner, 1992). Members of the immunoglobulin superfamily mediate Ca^{2+}-independent cell-cell adhesion either by heterophilic interaction or, in some cases, both homophilically and heterophilically (Cole et al., 1986; Cunningham et al., 1987). The members of the cadherin superfamily are the major CAMs responsible for Ca^{2+}-dependent homophilic cell-cell adhesion in vertebrate tissues (Geiger and Ayalon, 1992; Kemler et al., 1989; Takeichi et al., 1990).

1.2 The cadherin superfamily

The cadherin superfamily comprises a large number of cell adhesion molecules mediating Ca^{2+}-dependent adhesion of adjacent cells in a homophilic manner (Geiger and Ayalon, 1992). The members are surface glycoproteins ranging in mass between 120-140 kD. Cadherins are characterized by a N-terminal extracellular portion composed of tandemly repeated domains of approximately 110 amino acid residues (“cadherin repeats”), a single hydrophobic transmembrane region, and a cytoplasmic tail. They play a key role as morphogenetic regulators during embryogenesis (Huber et al., 1996), in the formation of junctional complexes and in the induction of the polarized cell type (McNeill et al., 1990), as well as in developing physical cell-cell associations (Angres et al., 1996). Distinct members of the cadherin superfamily also function as tumor invasion suppressors (Birchmeier and Behrens, 1994). Expression of each of the different cadherins is strictly regulated. Since most cell types are able to express more than one cadherin, characteristic patterns of tissue and cell distribution emerge. Reflecting their role as morphogenetic regulators, the expression patterns for cadherins correlate with steps of cell differentiation and tissue morphogenesis.

The increasing number of cadherins identified during recent years has been classified into subfamilies according to their structural features (Pouliot et al., 1992). Classical cadherins are characterized by their highly conserved cytoplasmic domains and the presence of five homologous repeats in the extracellular region of the molecule designated EC1-EC5

beginning with the N-terminus of the molecule (Kemler et al., 1992). The conserved cytoplasmic tail of these cadherins interacts indirectly with actin filaments by means of a group of intracellular proteins called catenins, including α-catenin, β-catenin, plakoglobin (γ-catenin), and p120 (Aberle et al., 1996; Geiger and Ayalon, 1992). A more detailed description of the classical cadherins is given below.

Desmosomal cadherins represent another cadherin family (Koch et al., 1994). The structure of the two members belonging to this family, desmocollins and desmogleins, resembles the structure of classical cadherins in their extracellular region, whereas their cytoplasmic domains differ significantly. Desmogleins have a C-terminal cytoplasmic domain containing, in addition to the cadherin-type segment, a number of repeats of a 29 ± 1 residue sequence not present in other cadherins. The members of this family interact with intermediate filaments rather than with actin filaments. Their cytoplasmic domain binds to a different set of intracellular proteins, which in turn bind to intermediate filaments. Thus, through desmosomes, the intermediate filaments of adjacent cells are linked into a net that extends throughout the many cells of a tissue.

Another cadherin family, glycosylphosphatidylinositol (GPI)-anchored cadherins, exhibits an extracellular structure similar to that of classical cadherins, but is entirely lacking the transmembrane and cytoplasmic domains. The extracellular portion is attached to the outer membrane surface by a glycosylphosphatidylinositol (GPI) anchor. T-cadherin from chicken and its human analogue, cadherin-13, belong to the GPI-anchored cadherins (Ranscht and Dours-Zimmermann, 1991; Tanihara et al., 1994).

Protocadherins are found in the central nervous system in the brain. They are distinguished from classical cadherins both by one or two additional extracellular cadherin-type repeats and by a complete lack of homology of their cytoplasmic domain (Shapiro et al., 1995). Their extracellular domains have several common structural features.

The 7D-cadherins belong to a family distinct from other cadherins by structural and functional features. The members of this family are composed of seven instead of five extracellular cadherin repeats. In addition, a much shorter domain consisting of only 21 amino acids replaces the highly conserved cytoplasmic region of the classical cadherins. Representatives of this family include LI-cadherin (Berndorff et al., 1994) and Ksp-cadherin (Thomson et al., 1995). The major structural features of members of the cadherin superfamily are schematically represented in figure 1.1.

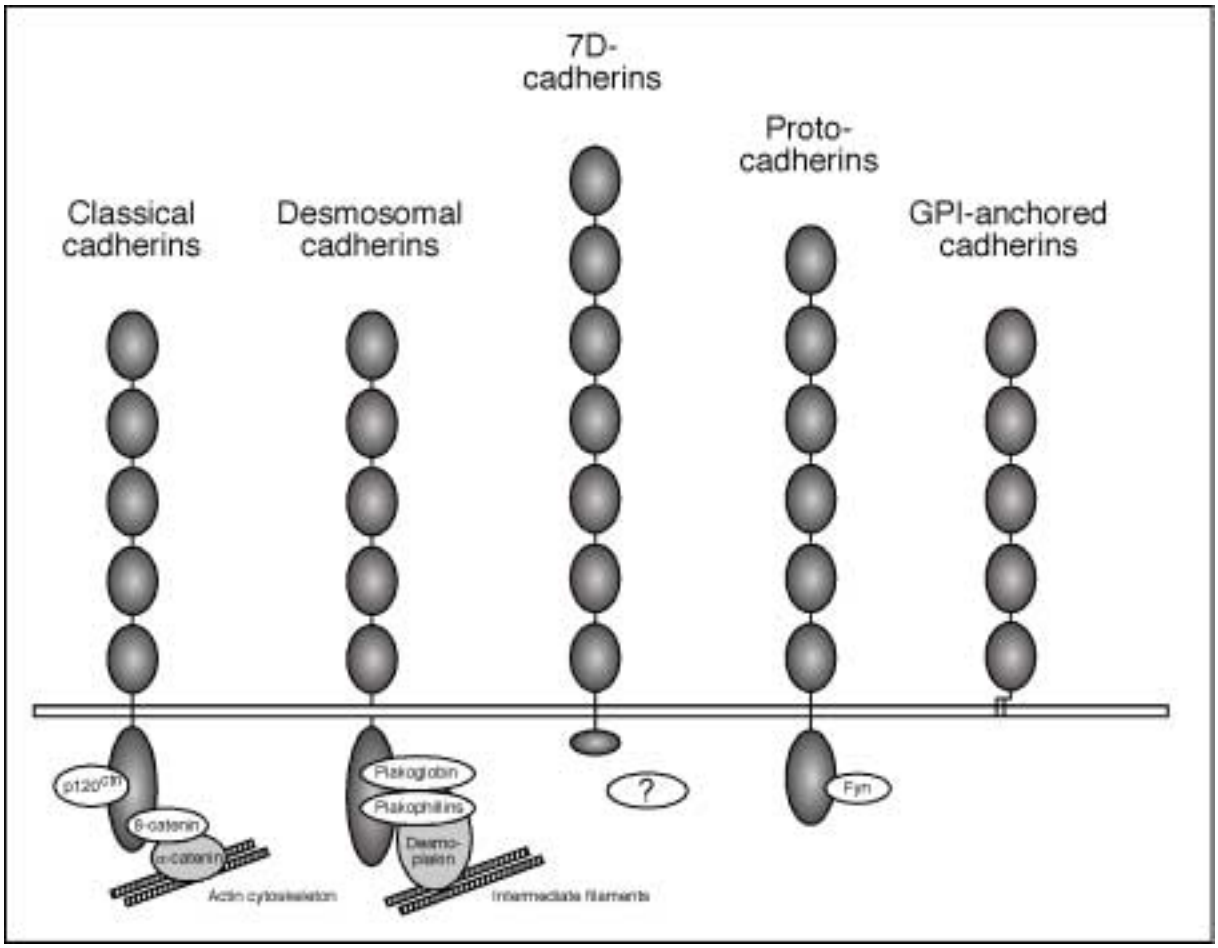

Figure 1.1: Schematic representation of the major structural features of the members of the cadherin superfamily.

1.2.1 Classical cadherin

The classical cadherins are transmembrane glycoproteins that exhibit, in addition to the structural homologies, a very similar overall protein topology. Protein sequence comparison has led to the identification of domains of common functional importance; the sequence His-Ala-Val located in the C-terminal region of EC1 plays a key role in the interaction between cadherins (Blaschuk et al., 1990). However, homophilic recognition also requires the presence of other regions located in the N-terminal domain. In addition, the conformation of the cadherin molecule is stable only in the presence of Ca^{2+}, whose binding is a prerequisite for cadherin-mediated cell-cell adhesion. Calcium-binding sites consist of short, highly conserved amino acid sequences located between neighboring extracellular repeats (Ozawa et al., 1990).

The cytoplasmic domains of classical cadherins associate with peripheral cytoplasmic proteins termed α-catenin, β-catenin, plakoglobin (γ-catenin), and p120 (Aberle et al., 1996; Geiger and Ayalon, 1992). This complex formation serves as intermediate link between the cadherins and actin filaments and regulates the adhesive function of cadherins (Ozawa et al., 1990).

E-cadherin (Kemler et al., 1977), N-cadherin (Hatta and Takeichi, 1986), and P-cadherin (Nose et al., 1987) are the most intensively studied representatives of the classical cadherins.

They are named according to the main tissue in which they were found: E-cadherin is present in many types of epithelial cells; N-cadherin in muscle, and lens; and P-cadherin in placental and epidermal cells. All have by now also been discovered in various other tissues.

E-cadherin is concentrated in adherens junctions in mature epithelial cells, where it helps to connect the actin cytoskeleton of the cells it holds together. It is also the first cadherin expressed during mammalian development (Ohsugi et al., 1996); E-cadherin expression is down-regulated during epithelial-mesenchymal transitions and reappears during reversion to the epithelial phenotype. E-cadherin is thus a faithful component in all epithelia and plays a functional role that is essential also for the maintenance of the epithelial phenotype (Behrens et al., 1989; Imhof et al., 1983). Alterations in E-cadherin expression are associated with tissue invasiveness, cellular dedifferentiation, and hyperplastic cell growth (Geiger and Ayalon, 1992).

P-cadherin is expressed in the placenta and in stratified squamous epithelia, and it is associated with an immature, proliferative cellular phenotype (Hirai et al., 1989). Similar to E-cadherin, P-cadherin is located in adherens junctions (Shimoyama et al., 1989). Despite its absence from the normal colon, aberrant expression of P-cadherin has been shown to be intimately associated with altered colorectal tissue architecture in colon cancer (Haq et al., 2001).

Neural (N)-cadherin is present in significant amounts in the endothelium, but in contrast to E- and P-cadherin, N-cadherin is not clustered at cell-cell junctions but remains diffuse on the cell membrane (Salomon et al., 1992). The lack of junctional distribution of N-cadherin is due to competition with vascular endothelial (VE) cadherin, which is another major cadherin expressed in endothelial cells (Navarro et al., 1998). However, when the clustering of VE-cadherin cannot occur for lack of homotypic interaction, N-cadherin retains full capacity to localize at cell-cell contacts (Navarro et al., 1998). Thus, cadherin localization at cell-cell junctions can be effected by other co-expressed cadherins rather than by cell type specific effects.

1.2.2 7D-cadherins

The first member of the 7D-cadherin family was detected in the liver and intestine in rat, and was thus assigned the name liver-intestine (LI) cadherin (Berndorff et al., 1994). In contrast, the homologous molecule in mouse and human is only expressed in polarized epithelia of intestine (Zitt, 1997). LI-cadherin is distinguished from the other cadherins by structural differences in the extracellular as well as in the cytoplasmic portion. The extracellular portion of LI-cadherin consists of seven instead of five structurally defined cadherin repeats (EC1-EC7), and the cytoplasmic domain has only 21 amino acid residues that show no significant homology to classical cadherins. In intestinal epithelial cells, LI-cadherin is evenly distributed over the lateral surface but is excluded from adherens junctions (Zitt, 1997), whereas co-expressed E-cadherin is concentrated in this specialized membrane region (Hermiston and Gordon, 1995). The cytoplasmic tail of LI-cadherin is neither associated with catenins nor is it tightly connected to the cytoskeleton. Nevertheless, LI-cadherin is able to mediate Ca^{2+}-dependent cell-cell adhesion (Kreft et al., 1997).

A protein with the same overall structure was found to be expressed exclusively in the basolateral membrane of tubular epithelial cells in the kidney, and was thus named kidney-specific (Ksp) cadherin (Thomson et al., 1995). The function of Ksp-cadherin is unknown, but it exhibits the Ca^{2+}-dependent sensitivity to proteolysis that is typical for other cadherin family members (Thomson et al., 1995).

1.3 LI-cadherin

LI-cadherin is selectively expressed in the basolateral surface of polarized hepatocytes and intestinal cells (Berndorff et al., 1994). In mouse adult tissue, it is expressed in small amounts in the spleen as well. In contrast to rat, mouse LI-cadherin is not expressed in the liver (Angres et al., 2001).

LI-cadherin is co-expressed with E-cadherin. However, while E-cadherin is found to accumulate in adherens junctions, LI-cadherin is distributed over the entire basolateral cell surface of neighboring enterocytes, goblet cells and paneth cells (Angres et al., 2001), but is always excluded from adherens junctions. All these cells develop from a stem cell compartment located in the crypts. Stem cells proliferate and the progenies differentiate into enterocytes and goblet cells (moving upward along the crypt villus axis) and into paneth cells (remaining at the bottom of the crypt). Within about five days, the enterocytes reach the top of the villus, become apoptopic and are extruded into the lumen of the intestine. In figure 1.2, a schematic representation of the intestinal epithelium is shown on the left. In the right panel, a section of a mouse small intestine is stained with anti LI-cadherin antibodies; it can clearly be seen that the expression of LI-cadherin is constant along the crypt-villus axis.

The production of the distinct cell lineages from a fixed stem cell population represents a complex equilibrium that undergoes continual renewal while maintaining precise interrelations. Whereas E-cadherin is essential for the maintenance of epithelial integrity (Behrens et al., 1989; Imhof et al., 1983) as well as for control of tissue differentiation (Larue et al., 1996), the physiological function of LI-cadherin still needs to be explored in the intact organism. Knockout studies are under way to analyze its biological function in detail.

During fetal development, the mammalian intestine undergoes considerable change in morphology and cell differentiation; the undifferentiated epithelium develops into a single-layered columnar epithelium, and villi are formed. These morphological changes occur in the mouse embryo between days 14.5 and 16.5. The expression pattern of LI-cadherin during mouse embryogenesis has shown that LI-cadherin is expressed on embryonic day 12.5 (Angres et al., 2001). Interestingly, LI-cadherin expression coincides with the morphological transformation of the multilayered epithelium into a single-layered columnar epithelium, a developmental stage during which major tissue remodeling, growth, and differentiation occur in the embryonic intestine. However, the expression of LI-cadherin does not correlate with the separation or fusion of tissues. This observation indicates that LI-cadherin does not participate in morphogenesis by a differential adhesion mechanism, as has been described for E-cadherin involved in early development (Takeichi, 1987). Whether LI-cadherin has a part in the development is unknown to date.

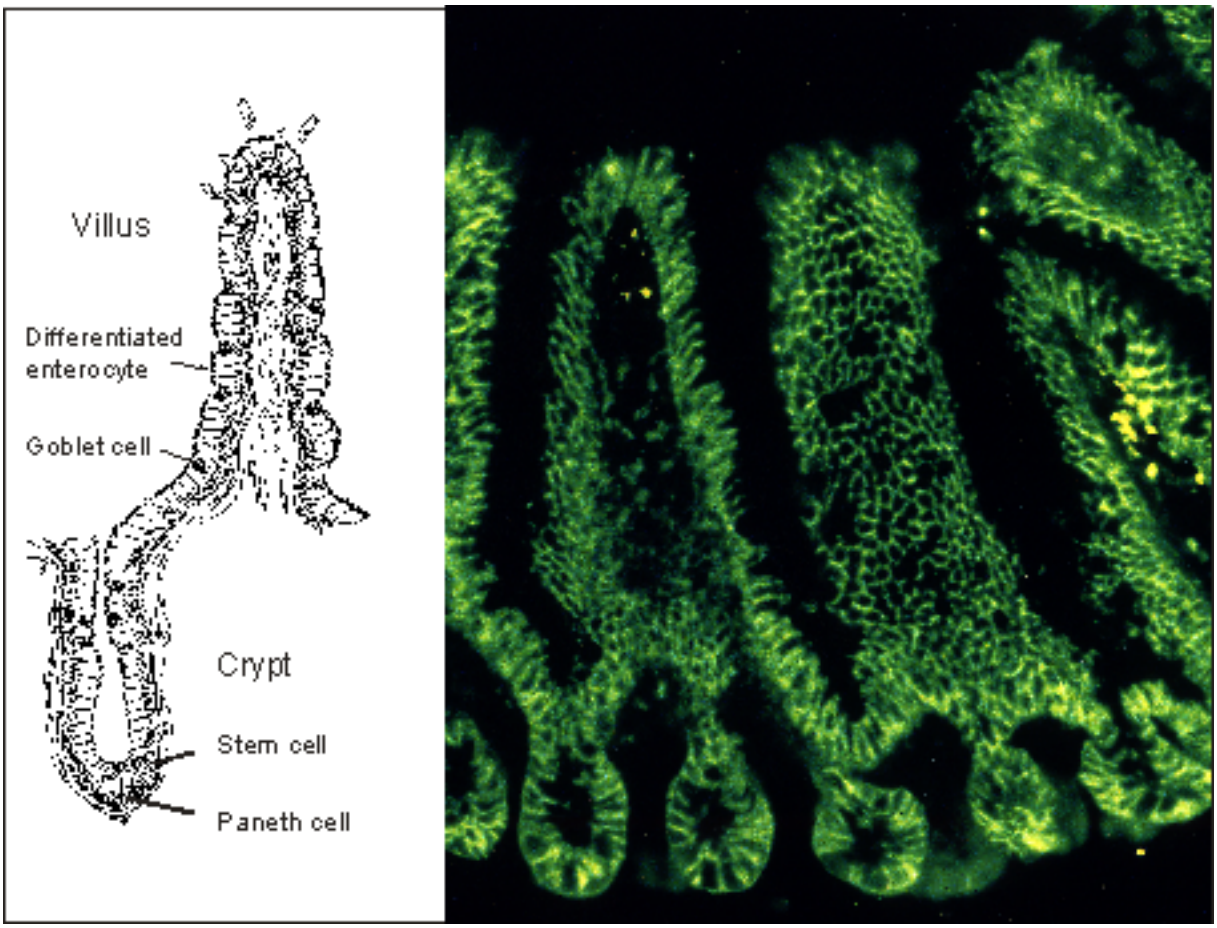

Figure 1.2: *Left:* Schematic view of the intestinal epithelium. *Right:* A section of small intestine (mouse) stained with anti LI-cadherin antibodies and FITC-labeled secondary antibodies.

Recently, LI-cadherin has been detected in metaplastic mucosa of the stomach as well as in gastric adenocarcinomas of the intestinal type (Grotzinger et al., 2001). While cadherins are known to be morphoregulatory proteins (Takeichi, 1995), it is conceivable that LI-cadherin is involved in these cellular transitions. Induction of LI-cadherin expression during intestinal metaplasia is paralleled by the observation that induced P-cadherin expression may participate in the morphogenesis of well-differentiated adenocarcinomas (Yasui et al., 1993). LI-cadherin may therefore become an important marker for gastric metaplasia and neoplasia.

1.4 Regulation of the expression of classical cadherins and 7D-cadherins

During embryonic development, cadherin expression occurs very early, and an exact spatio-temporal program for the different members of this protein family is essential for a correct morphogenesis of the embryo (Takeichi, 1988). In the adult organism, the specific expression of cadherins plays an important role in the maintenance of tissue architecture. During tumorigenic processes, misexpression of cadherins occurs and may explain the disruption of cell-cell associations that frequently take place in tumor cells. Specifically, down-regulation of E-cadherin has been reported in a high variety of human tumors and in animal models (Birchmeier and Behrens, 1994; Takeichi, 1993). In general, expression of E-cadherin in carcinomas is strongly correlated with the differentiation grade of the tumors. In contrast, increased expression of P-cadherin is observed in papillomas and squamous cell carcinomas.

Both molecules are completely suppressed in fully undifferentiated spindle cell carcinomas (Cano et al., 1996).

In order to understand the regulation of cadherin expression during normal development and neoplasia, it is important to study the regulatory elements that control cadherin gene expression. E-cadherin is expressed in all embryonal and adult epithelial tissues. Epithelium-specific expression of the E-cadherin gene is achieved by two different mechanisms. A 12-bp palindromic element (E-Pal) located in the upstream promoter suppresses transcription in mesenchymal cells, while in epithelial cells E-Pal activates transcription together with an epithelium-specific enhancer (ESE) located in the first intron of the gene (Hennig et al., 1996). In addition, a GC-rich region and a CCAAT box are identified as positive regulatory elements in the 5' proximal sequence.

The 5' proximal P-cadherin promoter region has some similarities to the mouse E-cadherin promoter, such as a CCAAT box and a GC-rich region, but they differ in the absence of a palindromic element (E-pal) in the P-cadherin promoter (Faraldo et al., 1993). The CCAAT-box and the GC-rich region have been shown to be involved in cell-type specificity and optimal promoter activity (Faraldo et al., 1997). Modification of some of the factors binding to the CCAAT-box and the GC-rich region in the P- and E-cadherin promoter, or alternation in their interaction with co-activators, seem to be involved in the down-regulation of E- and P-cadherin expression in carcinoma-derived keratinocyte cell lines.

Kidney-specific cadherin (Ksp-cadherin, cadherin 16) is a tissue-specific member of the 7D-cadherin family that is expressed exclusively in the basolateral membrane of tubular epithelial cells in the kidney. Similar to the E- and P-cadherin promoter, the Ksp-cadherin promoter is TATA-less and contains GC-rich regions and a CCAAT box. In addition, several consensus binding sites for transcription factors that mediate tissue-specific gene expression have been identified, including activator protein-2 (AP-2), hepatocyte nuclear factor-1 (HNF-1α and HNF-1β), CCAAT/enhancer-binding protein (C/EBP), basic helix-loop-helix (bHLH) proteins, and GATA factors. Recently, a 82-bp region containing the consensus HNF-1 site has been shown to be essential for activity in transfected renal epithelial cells. Both HNF-1α and HNF-1β bind to and regulate the promoter (Bai et al., 2002). However, 1342 bp of the 5'-flanking region are necessary for recapitulating the complete expression pattern of the Ksp-cadherin gene in transgenic mice, whereas transgenes containing 324 bp of the 5'-flanking region exhibit variegating effects but are also expressed exclusively in the kidney (Shao et al., 2002).

1.5 Gene specific expression in intestinal epithelial cells

During development and in the adult tissue the cellular phenotypes are defined by the expression of specific genes in the individual cells. The sets of genes expressed are principally determined by transcriptional initiation, and the particular set of genes expressed in a single cell type is referred to as the "transcriptome". Intestinal epithelial cell transcriptomes shift in well-orchestrated patterns during development, differentiation, and adaptive processes in the intestinal mucosa. Understanding the molecular mechanism that regulates transcription of the cellular gene sets is the foundation for understanding development and differentiation events.

Current understanding of transcriptional regulatory mechanisms in the intestine is based on the analysis of a few genes expressed in restricted cellular and developmental patterns. The gene encoding the disaccharidase sucrase isomaltase (SI) has many attributes that makes it a good choice as a model intestinal gene for the investigation of the mechanisms that direct intestine-specific gene expression (Traber and Silberg, 1996). Its expression is limited to the cells of the small intestine enterocytic lineage. In addition, it has a complex pattern of expression along the small intestinal crypt-villus axis, which follows the general pattern of development of enterocytes as they migrate from the crypt to the villus.

A number of transcription factors vital for SI gene transcription have been characterized, including hepatocyte nuclear factor-1 (HNF-1) (Wu et al., 1994), caudal-related homeodomain proteins (Cdx) (Suh et al., 1994), and GATA-4 (Boudreau et al., 2002). These factors activate the SI promoter activity in co-transfection experiments where GATA-4 requires the presence of both HNF-1α and Cdx2 to influence this activity (Boudreau et al., 2002). Thus, the temporal and spatial patterns of *SI* gene transcription in the intestinal epithelium depend on coordinated interactions of different DNA-binding proteins. Modulation of these complex DNA-protein and protein-protein interactions may occur via phosphorylation of specific proteins. Protein modifications are probably linked to cellular signaling processes that are activated by extracellular cues such as matrix, growth factors or adjacent cells. This shows how complex the regulating network of gene transcription is, and many processes that extend beyond simple expression of transcription factors modulate that transcription.

While expression of sucrase isomaltase is limited to the small intestinal crypt-villus axis, it serves as a marker protein for the intestinally differentiated epithelium. Villin is a second marker protein expressed within the intestinal crypts in the proliferative stem cells (Boller et al., 1988; Robine et al., 1985). It is located in the core of actin filaments supporting the microvillus membrane forming the brush border (Bretscher and Weber, 1979). In neoplastic tissues, villin expression is retained in most tumors or cell lines derived from villin-positive epithelia, but has never been observed in non-epithelial tumors (Moll et al., 1987). This pattern of villin synthesis in normal adult tissues and carcinomas has practical consequences in the differential diagnosis of human tumors. Interestingly, increased expression of villin coincides with induction of LI-cadherin expression in intestinal metaplasia and in gastric adenocarcinomas (Grotzinger et al., 2001).

Other genes specifically expressed in the small intestine include intestinal fatty acid binding protein (I-FABP) and lactase-phlorizin hydrolase (LPH). I-FABP is believed to participate in the uptake, intracellular metabolism, and transport of long chain fatty acids within enterocytes. Two regions among the 260 nucleotides immediately 5' to the start site of transcription have been found to be necessary for establishing and maintaining its region-specific expression along the duodenal-colonic axis of the perpetually renewing gut epithelium (Green et al., 1992). LPH, the enzyme critical for the digestion of milk lactose, is an absorptive enterocyte-specific protein that serves as a marker for intestine-specific gene expression and intestinal differentiation (Krasinski et al., 1994). GATA, HNF-1α, and Cdx2 are activators of the LPH promoter, and it has been shown that physical association between

GATA-5 and HNF-1α as well as Cdx2 and HNF-1α is required for the synergistic activation of the LPH promoter (Mitchelmore et al., 2000; van Wering et al., 2002).

In the intestinal mucosal epithelium, other members of the cadherin family than LI-cadherin are also represented. E-cadherin is detected in the enterocytes along the entire length of the crypt-villus axis (Hermiston and Gordon, 1995). In addition to E-cadherin and LI-cadherin, a third cadherin, R-cadherin, is also expressed in fetal intestinal epithelial cells (Sjodin et al., 1995). It remains to be explored why different types of cadherins occur simultaneously in the intestinal epithelium.

1.6 Aim

The LI-cadherin (cadherin-17) is a structurally distinct member of the cadherin superfamily with seven extracellular cadherin repeats, exclusively expressed in the small and large intestine but not in the upper gastric tract. Although LI-cadherin is absent from the healthy gastric epithelia, it becomes strongly expressed during intestinal metaplasia and in gastric adenocarcinomas. The onset of LI-cadherin expression during the murine embryonic development coincides with the morphological transformation of the multilayered epithelium into a single-layered columnar epithelium, a development stage during which major tissue remodeling, growth, and differentiation occur in the embryonic intestine.

The goal of this study was to analyze the LI-cadherin promoter region in order to reveal the transcriptional regulation of LI-cadherin gene expression. To achieve this, the following experimental tasks had to be performed:

1. Identification of the transcription start site.
2. Analysis of the transcriptional activity of the assumed promoter region in different cell lines.
3. Identification of transcription factor binding sites within the promoter region.
4. Construction of serial deletion mutants and quantification of their transcriptional activity in tissue specific cell lines.
5. Functional definition of transcription factor binding sites by perfoming electrophoretic mobility shift assays (EMSA).
6. Mutational analysis of the identified transcription factor binding sites in transfected cells in order to confirm their regulatory effect.

It is anticipated that these studies will reveal which control elements or regulatory DNA-binding proteins may be responsible for the tissue specific expression pattern of LI-cadherin. The analysis may also provide insight into the role of LI-cadherin for the cellular differentiation process during embryonic development.