ANALYSIS OF AN EVOLUTIONARILY CONSERVED DISTANT REGULATORY REGION DOWNSTREAM OF THE PAX6 GENE

by

JÖRN LAKOWSKI

(UNDER THE DIRECTION OF JAMES LAUDERDALE)

ABSTRACT

The evolutionarily conserved transcription factor Pax6 is crucial to the development of a variety of tissues including the central nervous system, the eyes and endocrine glands. It controls the expression of wide variety of factors including other transcription factors, cell-cell signaling molecules and cell adhesion molecules, thereby regulating important developmental processes such as cell proliferation, cell adhesion, migration and cell-signaling. Although, many of the downstream targets of Pax6 are known, the upstream regulators of *Pax6* remain largely unknown. In this study, we have investigated the role and mechanism of a novel regulatory region (DRR), located 150 kb downstream to the Pax6 transcription unit. Analysis of the regulatory region indicates, that it is located in the intron of the housekeeping gene *Elp4*, which we show, is regulated by a TATA-less bidirectional promoter. Within the DRR, we have identified an enhancer element, which we propose acts on *Pax6* but not on *Elp4* due to promoter specificity.

INDEX WORDS*: Pax6, Elp4,* Bidirectional promoter, Gene regulation

ANALYSIS OF AN EVOLUTIONARILY CONSERVED DISTANT REGULATORY REGION DOWNSTREAM OF THE *PAX6* GENE

by

JÖRN LAKOWSKI

VORDIPLOM, UNIVERSITY OF ROSTOCK, GERMANY, 1999

A Thesis submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirement for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2003

ANALYSIS OF AN EVOLUTIONARILY CONSERVED, DISTANT REGULATORY REGION DOWNSTREAM OF THE *PAX6* GENE

by

JÖRN LAKOWSKI

Major Professor: James Lauderdale

Committee: Nancy Manley
Marcus Fechheimer

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2003

TABLE OF CONTENTS

PAGE

LIST OF FIGURES

CHAPTER 1. BACKGROUND

1.1. OVERVIEW OF MOUSE EYE MORPHOGENESIS

Mammalian eyes are generated as a consequence of a series of inductive signals and regional specification events, that are first evident as an outpocketing of the forebrain neuroectoderm and the overlying surface ectoderm (Fig. 1)[1, 2]. The developmental process can be subdivided into distinct steps. The first event to take place is the definition of the eye field in the anterior neural plate, which then evaginates from the forebrain to become the optic pit and subsequently the optic vesicle. The optic vesicle then further evaginates toward the surface ectoderm, which has latent lens forming abilities. Upon contact, the optic cup triggers the formation of the lens placode from the overlying surface ectoderm. The lens vesicle then invaginates and pinches off from the ectoderm to form the lens vesicle and subsequently the lens. The cornea will be formed from the remaining placode. At the same time, the medial part of the optic vesicle narrows to form the optic stalk, and the proximal part is pushed in, due to the lens placode invagination, to generate the two-layered optic cup whose two layers will differentiate in different ways. The outer layer of the optic cup will develop into the melanin expressing retinal pigmented epithelium. The cells of the inner layer will divide rapidly and generate a large variety of cells, which constitute the neural retina, such as light-sensitive photoreceptors, ganglion cells and Mueller glia cells. Of these various cell types, the ganglion cells are the ones whose axons send out electrical impulses to the brain in order to transmit visual

Figure 1.

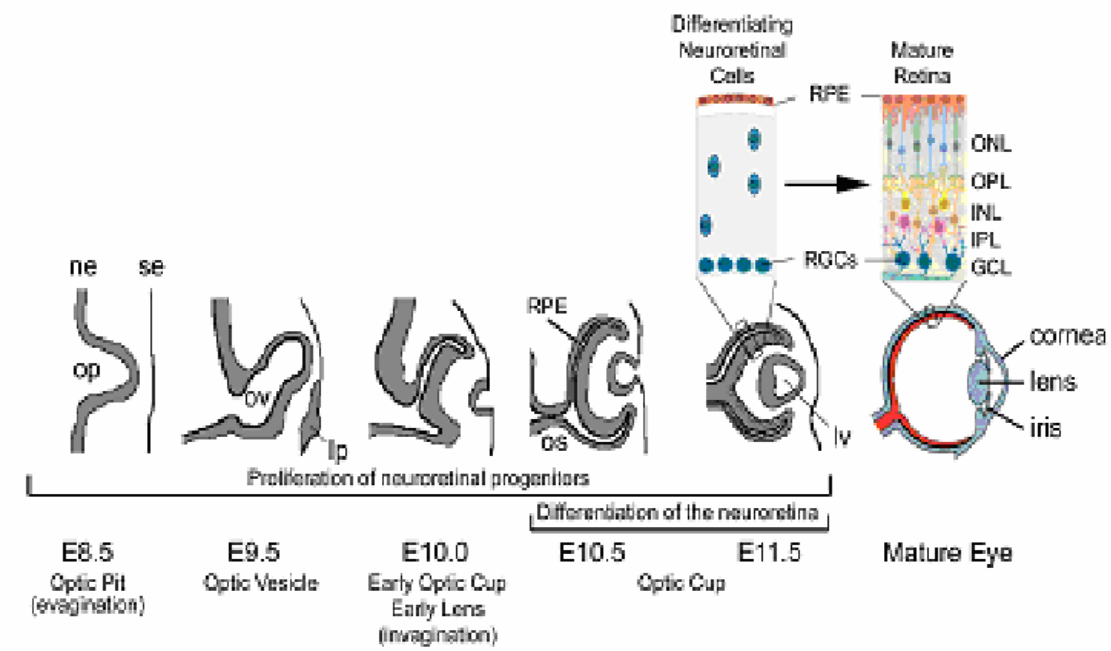


Figure 1. **Mouse eye morphogenesis from embryonic day 8.5 to 11.5**. Optic vesicle is formed around day 8.5 of embryonic development as an evagination of the neural ectoderm. Upon contact with the surface ectoderm the lens placode is induced which subsequently pinches off to give rise to the lens vesicle. The center of the optic vesicle then invaginates to form the bi-layered optic cup. Inset shows the development of the multilayered neural retina from undifferentiated retinal progenitor cells to mature neurons. neural ectoderm; op, optic pit; ov, optic vesicle; lp, lens placode; RPE, retinal pigmented epithelium; os, optic stalk, lv, lens vesicle; se, surface ectoderm

information. These axons meet at the base of the eye and travel down the optic stalk, which will eventually, gives rise to the optic nerve. The ciliary body and the iris develop from the region on the outer lip of the optic cup, where the prospective neural and pigmented retinal layers meet.

1.1.1. RETINOGENESIS

Because of its accessibility, the differentiation of the neural retina is a process, which is well studied and is a model example of induction and differentiation processes. During retinal development, postmitotic cells are generated in the germinative layer and radially, forming the laminar pattern of the retina. The neural retina develops into an array of different cell types, which all originate from a common population of multipotent retinal progenitor cells residing in the inner layer of the optic cup.

The cell types generated include the light-sensitive rod and cone photoreceptors, bipolar interneurons transmitting electric stimuli, amacrine and horizontal neurons, and the ganglion cells. In addition to these cell types, the retinal precursor cells also give rise to Mueller glia cells, which maintain the integrity of the retina. The retinal neurogenesis proceeds in a predictable and evolutionarily conserved, histogenetic order. The retinal ganglion cells and horizontal cells appear first, followed by the cone-photoreceptors, amacrine cells, rod-photoreceptors, bipolar cells and finally the Mueller glia cells.

1.1.2. LENS AND CORNEA DEVELOPMENT

After induction by the optic vesicle, the lens placode rounds up and eventually pinches off to form the lens vesicle, which now contacts the newly formed overlying ectoderm.

The lens vesicle then induces the ectoderm to form the transparent cornea. The process of corneal formation depends very much on physical factors. For example, the correct intraocular fluid pressure is necessary to achieve the necessary corneal curvature to appropriately focus the light on the retina.

In order to form a functional lens, the lens tissue has to undergo specific changes in cell structure and shape and is associated with the synthesis of lens-specific proteins called crystallins, which are the major structural component of the lens.

The development of the lens vesicle differs regionally. The cells at the inner portion of the lens, facing the neural retina, differentiate into primary lens fiber cells. These cells elongate, synthesize crystallins and eventually loose their nucleus. The cells on the anterior side of the lens vesicle remain in an epithelial state and keep dividing. These dividing, anterior cells move toward the equator of the lens vesicle and start elongating upon passing the equatorial zone. In summary, the lens encompasses three different regions: an anterior region of dividing cells, a middle region of cellular elongation and a posterior zone of crystalline production. This pattern in maintained throughout life, as new fibers are continuously laid out.

1.1.3. DEVELOPMENT OF OTHER EYE STRUCTURES

Other eye structures besides the ones mentioned above include the iris, the ciliary body and the sclera. The iris is a pigmented, muscular tissue, which is unique to the vertebrate lineage. It is composed of the two ectodermal layers forming the edge of the optic cup and a layer of vascularized connective tissue, derived from surrounding mesenchyme. Its ability to constrict and dilate enables the individual to regulate the light influx and to

focus incoming light better on the neural retina. In contrast to other muscles, which are of mesodermal origin, part of the iris is derived from the ectoderm, specifically the outer rim of the optic cup. Furthermore, neural crest derived mesenchymal cells also populate the iris. The ciliary body is formed by the bilayered, optic cup, which undergoes folding to generate the ciliary processes. The mesenchyme at the edge of the optic cup gives rise to the connective tissue of the ciliary body, smooth muscles of the ciliary muscle and suspensory ligaments of the lens. The fibers, making up the suspensory ligaments, attach the ciliary process to the lens in a way that contraction of the ciliary muscles will result in an alteration of lens curvature and ultimately in focusing of light on the neural retina. The sclera is formed by the mesenchymal tissue surrounding the developing eye.

1.2. TRANSCRIPTION FACTOR PAX6

The Pax6 protein belongs to a family of transcriptional regulators which were first isolated through sequence homology to the DNA binding domain of the *Drosophila melanogaster* paired gene which is involved in the establishing body segmentation [3]. The proteins encoded by the *Pax* gene family all share this 128 amino acid DNA binding domain and play key roles in several developmental processes, particularly the nervous system [4, 5]. Of the nine family members, Pax6 has been studied most thoroughly in mice and human, partly due to the striking phenotypes associated with mutations in this gene and its paramount role in eye development. In addition, its role in eye development, Pax6 is required for the proper progression of a variety of developmental processes, particularly in the development of both sensory and nervous system in vertebrates and invertebrates. The Pax6 protein has several structural motifs. In addition to the paired

domain, Pax6 contains a 60 amino-acid homeodomain DNA binding motif and a proline-serine-threonine (PST) rich domain, which is thought to facilitates protein-protein interaction and has been shown to act as a transactivation domain [6]. The transcription unit of Pax6 extends approximately 23kb in humans and gives rise to a 2.7kb cDNA, which encompasses a large 5' and a 3' untranslated region with translation initiating in exon 4. Interestingly, the *Pax6* gene produces several different isoforms, with Pax6 and Pax6a being the most common ones. These two isoforms are generated by the alternative splicing of exon5a, and the inclusion of the alternative exon leads to the insertion of additional 14 amino acids into paired domain of the Pax6 protein which has been demonstrated to affect its DNA binding specificity for new target sites [7]. Specific knockout of the Pax6(5a) isoform leads to iris hypoplasia and retinal malformations but does not change other Pax6 expression domains indicating the importance of the alternative isoform for iris development [8]. Together with this observation and the fact that Pax6(5a) is conserved in the vertebrate lineage only, it was proposed that this isoform is crucial for the vertebrates to develop the iris as a structure to regulate the light influx into the eye and also to better focus the light onto the retina. Homolog of the Pax6 protein have been found and isolated from a vast variety of organisms including mouse, humans, flies and nematodes. The amino acid sequences, as well as the expression pattern of this protein are highly conserved throughout the vertebrates. In fact, the protein is so highly conserved that when mouse Pax6 is expressed ectopically in *Drosophila* it is capable of inducing complete fly eye structures, indicating that the molecular pathways of eye generation have also been conserved over such a tremendous evolutionary distance [2]. Because of it's crucial role in eye formation Pax6 is also often referred to as the

“master regulator“ of eye development and the degree of structural conservation argues for a functional significance of most amino-acid. The expression pattern of Pax6 is conserved and very distinct. In mice, transcripts are first detected in the anterior neuroectoderm on day 8 of embryonic development. On day 8.5, expression is detected in the region of the region that will eventually give rise to the optic cup [2, 9]. Later on in development, Pax6 expression is seen in a more restricted pattern. On day 10.5 Pax6 positive structures include the developing eyes, nose, diencephalon, rhombencephalon, neural tube and nasal placode (Fig.2). In contrast, expression in the adult is only maintained in the eye, cerebellum and the endocrine pancreas.

1.2.1. ROLE OF PAX6 IN EYE MORPHOGENESIS

Pax6 plays a pivotal role in eye induction and has additionally been implicated in a variety of events during eye organogenesis. Detailed analysis of its cellular functions suggests involvement in processes such as cell proliferation, differentiation and adhesion/migration [10]. Although, Pax6 null mutants in mice initiate the formation of the optic vesicle and induction of the lens placode, as measured by Sox2 expression, lens development arrests and both lens and retinal structures subsequently degenerate. It is thought, that the optic vesicle induces the lens placode by secretion of a molecule(s) and that Pax6 confers the ability to respond to this secreted factor. Furthermore, a reduced proliferative capability in the developing lens has been shown for heterozygous mutant cells. Pax6 protein also plays an important role in the induction of certain crystallins in the forming lens at the stage of placode and lens vesicle. It was demonstrated that Pax6 has the capability of binding to the promoter region of the alpha-A-crystallin gene and

Figure 2.

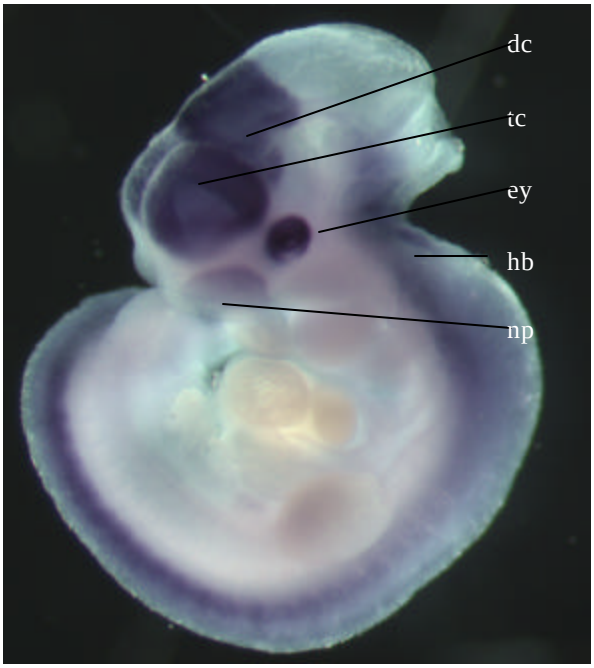


Figure 2. **Expression pattern of Pax6**. Whole mount *in situ* hybridization of Pax6 in a mouse embryo at E10.5 of development (lateral view). Distinct expession is seen in telencephalon (tc), diencephalons (dc), eye (ey), hindbrain (hb) and along the neural tube (nt).

alpha-B-crystallin gene in mouse, and several other types of crystallin gene in chick and guinea pig. In addition, ectopic expression of Pax6 in *Xenopus leavis* leads to the formation of lens-like structures. These and other findings suggest a major role for Pax6 in the regulation of crystallin genes during development. Although, Pax6 is dispensable for optic vesicle formation, it does play a role in subsequent steps of retinogenesis. At the optic cup stage, it is required for cell proliferation and differentiation [9]. Together with other factors such as Math1 and Math5, Pax6 is thought to maintain the multipotency and proliferation of the retinal progenitor cells and is required for these cells to develop the full neuronal potential. Pax6 gene expression in the retina is maintained in all proliferating retinal progenitor cells, but becomes downregulated after the differentiation of most neuronal cell types, except for bipolar, ganglion and amacrine neurons.

1.2.2. ROLE OF PAX6 IN OTHER DEVELOPMENTAL PROCESSES

Pax6 expression is not restricted to the developing eyes but is also observed in both neural and non-neural sites. In the early stages of development *Pax6* expression can be seen in the nasal structures, ductal and endocrine pancreas, pituitary, brain and spinal chord. Most of our current knowledge of Pax6 function comes from studies using the S*mall eye* mouse model, which is characterized by a reduced size of the eyes [11]. The most commonly used strains are $Pax6^{Sey}$ and $Pax6^{Sey\text{-}Neu}$ which have very similar phenotypes with both alleles containing a mutation that results in a premature stop-codon. Mice homozygous for these mutations die neonatal, lack eyes and nasal structures and exhibit severe abnormalities of the central nervous system, including gross defects of the forebrain. In both *small eye* and knockout mice, it was shown that Pax6 is involved in

pancreatic gene expression and development. Pax6 expression is first observed in the fore/midgut endoderm, which give rise to the pancreatic bud. Later, it is coexpressed with insulin and glucagon positive cells and at the time of birth, *Pax6* is expressed in all hormone-producing cells. Analysis of the mutant mice demonstrated that the number of hormone secreting cells was greatly reduced and the residual islet structures appeared unorganized [12]. Furthermore, the production of insulin and glucagon is reduced, a result which is expected since functional consensus Pax6 binding sites have been identified in the promoter regions of proglucagon, somatostatin and insulin. However, the regulatory effects of Pax6 in hormonal gene expression is not restricted to the pancreas. It has become apparent that Pax6 may also influence gene regulation in the enteroendocrine cells of the large and small intestine as well as in the developing pituatary gland [13]. For example, $Pax6^{Sey}$ embryos have a much smaller number of growth hormone producing cells, and serum levels in mutants is only about 25% compared to wild type embryos.

In addition to the roles mentioned above, Pax6 appears to be involved in the patterning of the central nervous system and cell signaling events therein. For instance, *Pax6* is mainly expressed in the dorsal part of the forebrain. Loss of the protein from that region results in the loss of expression of certain bHLH and homeodomain containing transcription factors and ectopic expression of other factors in the same region [14-16]. These findings indicate that Pax6 is responsible for the activation of certain transcription factors (e.g. Ngn 1/2) as well as the repression (e.g. Gsh 1/2) other factors in its domain of expression. Furthermore, several lines of evidence suggest that Pax6 is also involved in the regulation of cell surface properties in its expression domain [17]. Though not fully understood, this would probably have a very broad effect on developmental processes such as the

restriction of cells to certain territories, cell migration and axon outgrowth and target finding, all of which were found to be abnormal in Pax6 deficient mice.

1.2.3 REGULATION OF PAX6 EXPRESSION

A number of cis-acting regulatory regions have been defined for Pax6 in a variety of organisms. The murine Pax6 is transcribed from three distinct promoters P0, P1 and P-alpha which were shown to be differentially active during forebrain development [18]. Several enhancer elements have been identified 5´, within and 3´ to the actual Pax6 transcription unit. The identified regulatory regions behave as discrete modules, that direct expression of Pax6 in a highly tissue and/or cell type specific manner. These regions are also evolutionary conserved, as demonstrated by sequence and functional comparison. Approaches involving the generation of transgenic animals have been particularly useful for the identification of novel regulatory elements. A 107 bp region, 4.6 kb upstream of the P0 promoter, was shown to direct the expression of a lacZ reporter gene in transgenic mice to the lens and cornea in the appropriate developmental context [19]. Comparison of this region between human, mouse and puffer fish showed high sequence identity and revealed potential homeobox-binding sites for Pax6 regulators. In the same study, a regulatory region capable of directing lacZ expression to both neural and pigmented retinal tissue was identified. This 530 bp element, (designated as *alpha enhancer*) was mapped to a location just downstream of the translational start site and drove expression of the reporter gene predominantly in the nasal and temporal part of the retina. This region also displayed a number of potential homeobox-binding sites, among which were binding sites of Pax2 and Msx1 transcription factors involved in crucial steps

of eye morphogenesis. In a different study, two more eye tissue enhancer elements where identified, one of which not only directed reporter gene expression to a subpopulation of retinal progenitor cells but also to certain pancreatic cell populations as well [20]**.** The second enhancer element was shown to be active in the retinal cone cell population, where Pax6 is believed to have a function in the differentiation process. Interestingly, when the murine retinal enhancer, residing in intron 4, was tested for activity in *Drosophila melanogaster*, it was found that it was capable of driving the reporter gene in a subset of eye structures. Vice versa, the fly *ey* (eyeless) enhancer element was also shown to be active in the mouse retina, lens and other part of the central nervous system. These findings clearly demonstrate that the pathway upstream of Pax6, leading to the generation of eye structures, is highly evolutionarily conserved.

In addition to these well-characterized upstream enhancer, promoter and intronic regulatory elements, a whole series of new downstream regulators have recently been discovered and have partially been assessed. Their existence was first deduced from aniridia (see below) associated chromosomal rearrangements which were located well outside the known *Pax6* transcription unit [21]. In these cases of aniridia, the Pax6 gene was shown to be unaltered, and the phenotype was is associated with a number of downstream chromosomal rearrangements including inversions and translocation and deletions (see below).

1.3. ANIRIDIA

Aniridia is a rare disorder of the eye with a population frequency of approximately 1:60,000 to 1:100,000 [22, 23]. The severity of phenotypes associated with aniridia can

vary tremendously and can include some of the following symptoms: iris hypoplasia or absence, corneal opacification, lens dislocation, foveal dysplasia, strabismus and nystagmus. Glaucoma may also be observed in some patient who, and if untreated, can destroy the remaining vision. In summary, aniridia patients and generally suffer from a poor visual acuity. About 60% of aniridia cases are familial, usually with a high degree of penetrance. The remaining 40% are caused by sporadic mutations and show no family history. Aniridia is typically caused by heterozygous null mutation within the *Pax6* gene or cytogenetic deletions of the human chromosome 11p13, encompassing the *Pax6* transcription unit. Thus, aniridia arises from *Pax6* haploinsufficieny [24]. Almost all mutations lead to premature protein translation termination, causing a loss of activity of one allele. In general, heterozygosity for a null allele of Pax6 causes aniridia whereas heterozygosity for missense allele usually produces milder phenotypes.

1.3.1. GENOMIC REARRANGEMENTS 3' TO PAX6 ARE ASSOCIATED WITH ANIRIDIA

As mentioned above, most aniridia cases are caused by alterations in the Pax6 transcription unit and result in a non-functional protein. In contrast to *Drosophila*, where regulatory alleles have been identified for the Pax6 homolog *eyeless*, regulatory alleles of Pax6 are rarely found in mammals. The first hint of a regulatory allele came from a study of two independent aniridia pedigrees in which certain chromosomal rearrangements 3' to the Pax6 poly-adenylation signal were observed, but that left the transcription unit intact. An isolated yeast artificial chromosome (YAC) clone encompassing the Pax6 gene indicated that the rearrangements were at least 85kb downstream of the gene. It was first

proposed that due to these rearrangements, the Pax6 gene was now situated in an inappropriate chromatin environment for normal expression, and this positional effect was the underlying mechanism for the disorder observed in these two pedigrees. Several other similar cases have been reported since this finding. For example, taking advantage of a large number of available aniridia patients, Lauderdale et al. reported two submicroscopic de novo deletions of chromosome 11p13, 3' to Pax6 that caused sporadic aniridia in unrelated patients. In both cases aniridia phenotype was indistinguishable from Pax6 mutation [25]. Furthermore, using a human-mouse somatic cell hybrid system, Lauderdale et al. showed that the *Pax6* gene was only transcribed from the intact chromosome and not from the chromosome carrying the genomic rearrangement. These observations suggested remote regulatory regions, such as enhancers or a locus control region, more than 100 kb 3' to the Pax6 gene, which is essential for correct Pax6 expression. Strong evidence arguing for a 3' regulatory region came from experiment with yeast artificial chromosomes. Schedl et al, introduced into transgenic mice a 420 kb yeast artificial chromosome (Y593) carrying the entire human Pax6 transcription unit and large flanking regions [24]. When crossed into the *Small eye* background, the human transgene both rescued the mutant phenotype and reconstituted the normal mouse *Pax6* expression pattern. When a smaller, 310 kb (Y589) yeast artificial chromosome, lacking the large 3' flanking region, was introduced, it both failed to rescue the small eye phenotype and reconstitute the normal mouse Pax6 expression pattern at normal levels [26]. Thus, it can be said, that the 22kb Pax6 transcription unit needs to be present in a much larger genomic context for correct developmental expression and the region between the telomeric endpoints of YAC 593 and Y589 in particular seems to be crucial.

These ingenious experiments implicated a small 40 kb region, located ~155 kb downstream of the major P1 promoter, as important for *Pax6* regulation. To identify potential regulatory elements, Kleinjan et al. undertook DNase I hypersensitivity assay. This assay identified eight hypersensitive sites in several human cell lines. Because DNase I hypersensitive sites are usually associated with transcription factor binding sites, investigators then used DNA fragments containing several different sites to generate ß-galactosidase reporter constructs. Using this approach and evolutionary sequence comparison two different regulatory regions were identified that drive expression in several Pax6 specific domains, including the retina, lens, parts of the fore-and hindbrain as well as the spinal cord. Because of the apparent presence of regulatory sequences between the telomeric endpoints of YAC 593 and Y589, this region was named "downstream regulatory region" (DRR) (Fig. 3). Interestingly, the analysis of the aniridia-associated breakpoints revealed the presence of a second gene, subsequently designated as *Elp4*, which is transcribed in 3'-5' orientation relative to the Pax6 transcription unit (tail to tail orientation) with transcription starting approximately 200 kb distal to the Pax6 poly-adenylation signal [27]. The DRR is located in the final intron of this neighboring gene. The gene itself was found to be disrupted in several clinical aniridia cases indicating a potential role in aniridia.

1.4. PURPOSE OF THE STUDY

The overall goal of this work was to further the understanding of the complex developmental gene regulation of the transcription factor Pax6, a crucial player of many

Figure 3.

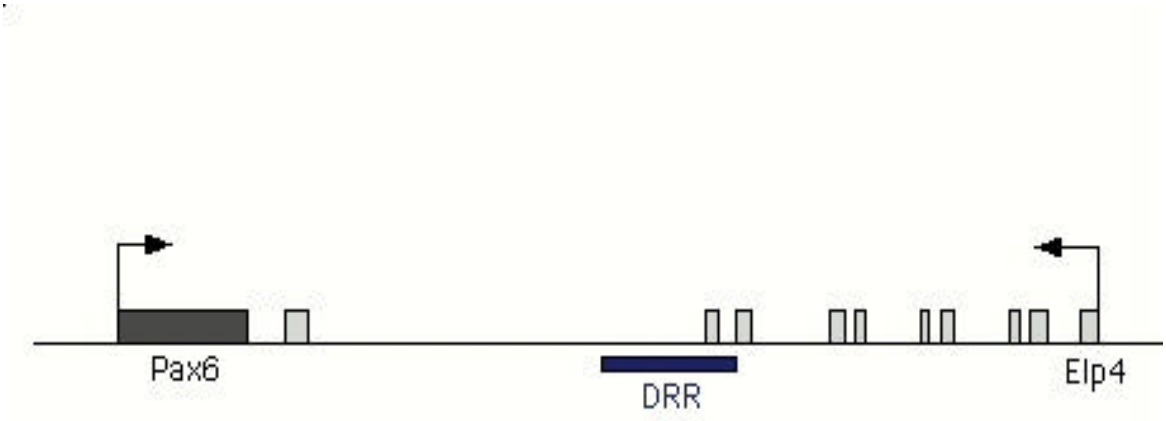


Figure 3. **Position of the downstream regulatory region (DRR) of Pax6 in the human genomic context of chromosome 11p13.** Spatial relationship between Pax6 and *Elp4* has been evolutionarily conserved with examples observed in rodents and fish. Pax6 transcription unit (one transcription initiation start site shown as arrow) shown as dark grey box on the left. *Elp4* transcription start site (arrow) and individual exons shown as grey boxes. Position of DRR is displayed as blue line centered between Pax6 and *Elp4*.

developmental processes. As summarized above, Pax6 has been the focus of many studies due to its importance during development. As a result, various downstream targets have been identified. The long list of factors that are targets of the Pax6 protein encompass cell-cell signaling molecules, cell adhesion molecules and other transcription factors. In contrast to that, very little is known about the factors and mechanisms controlling the expression and regulation of the Pax6 gene itself.

Most studies thus far have focused on regulatory sequences 5' to or within the Pax6 transcription unit. The discovery of chromosomal rearrangements at a long distance from the 3' end of Pax6 in patients suffering from the panocular disease aniridia and recent work in transgenic mice, prompted us to focus our attention on an approximately 40 kb region, now termed "downstream regulatory region" (DRR).

Interestingly, the regulatory region identified was found to be embedded in an intron of a second gene, transcribed in an antisense orientation relative to Pax6. Furthermore, it was shown that this neighboring gene was disrupted in several aniridic patients, invoking the possibility of a second aniridia gene. The genomic structure indicated that this DRR was located more than 100 kb downstream of the promoter regions of both genes, a very curious arrangement with regard to the gene regulatory mechanisms at work.

In order to further elucidate the role and mechanism of the DRR, we undertook a dual approach. First, we wanted to characterize the neighboring gene *Elp4* (formerly know as PAXNEB) as regards to its spatial and temporal expression pattern. We reasoned that if we found an eye specific expression pattern it would support the argument that this gene is indeed involved in aniridia. In addition, it is not very common, though not unprecedented, to find regulatory sequences of one gene in the intragenic region of

another gene. This alone created several questions as to what kind of transcriptional mechanism underlies the regulation of both genes with respect to the DRR and the chromatin environment.

Therefore, we decided to also analyze the *Elp4* promoter and upstream region, for one can deduce potential regulatory mechanisms from the structure and components of these elements. In a second approach, we tried to identify new regulatory sequences within the DRR in order to eventually be able to determine the transcription regulators binding to these sequences. We were able to identify several new elements using evolutionary sequence comparison (phylogenetic foot printing), which were then functionally tested in our *in vitro* system. Therefore, our goals can be summarized as follows; **1.** The Characterization of the *ELP4* expression pattern as well as analysis of regulatory mechanisms; **2.** Identification of novel regulatory sequences within the DRR and **3**. Propose a mechanistic model for the action of the DRR.

CHAPTER 2. MATERIALS AND METHODS

2.1. REPORTER PLASMID CONSTRUCTS

All reporter constructs were generated using standard molecular techniques unless indicated otherwise [28]. Pax6 enhancer constructs were made as follows.

Three of the four blocks of highly conserved sequence, found through evolutionary sequence comparison, were amplified using Pfx high-fidelity-DNA polymerase (Roche Molecular Biochemicals) using the following primers; Element B forward (5'CGTGCTTGGGTCTTATGGAG-3); reverse (5'-GACAAATACAAGTGTTTCAC-AGTGGT-3'); Element C forward (5'-GG-CAAGAATCTGAGGGTGT-3'); reverse (5'-CTCCTCTATTGCTTTGCTG-3'); Element D forward (5'-CCAACTTGAACCT-GCGAAG-3'); reverse (5'-GACGGTC-CAGGTAAGGGGAT-3'). PCR products were then cloned into pGL3-promoter vector (Promega) which had previously been modified for T-vector cloning according to a protocol by Marchuk et al. [29].

In addition to the SV40 promoter, we also tested the herpes simplex thimidine kinase promoter (HSV-TK) which is known to drive reporter genes at very low levels in mammalian cells. For that purpose, the SV40 promoter region was simply exchanged for the HSV-TK promoter using the intrinsic restrictions sites for Bgl II and Hind III contained within the pGL3 reporter vector.

For the *Elp4* promoter constructs, a 3.8kb Nco I fragment was subcloned into the promoter and enhancer-less pGL3-Basic vector creating *ELP4* 3.8 in two orientations.

A 550bp SacI-NcoI fragment was sub cloned into pGL3-Basic to generate *ELP4* SN. *ELP4* KK was made by cloning of a 300bp Kpn I fragment into pGL3-Basic in both orientations. An Nco I-Nhe I deletion fragment was generated by using the *ELP4* 3.8 giving rise to a 100bp test fragment. The construct *ELP4* SS was made through deletion of a Sac I-Sac I fragment from ELP 3.8 (3'-5' orientation).

2.2. DUAL LUCIFERASE ASSAY

To test the various reporter constructs we chose to use our established *in vitro* dual luciferase reporter system. This kind of reporter system is commonly used to improve the experimental accuracy and reduce the noise level inherent to cell culture assays. In this approach, two different, independent reporter enzymes are simultaneously expressed and measured in a single system. Usually, one reporter works as an internal control and establishes the baseline conditions, whereas the experimental reporter carries the cis-acting sequences to be tested for enhancer/promoter activity. The ratio between the values obtained from these two reporters represents the normalized assay outcome and helps to minimize experimental variability caused by differences in cell viability, transfection efficiency, pipetting volumes, cell lysis efficiency and assay efficiency. An experiment conducted this way allows a more reliable interpretation of the experimental data because extraneous influences are excluded. For our experiments, we used the DLR-Assay kit by Promega to obtain our data. The assay measures the activity of the Firefly (*Photinus pyralis*) and Renilla (*Renilla reniformis*) luciferases with readings taken sequentially. Because of the nature of the enzymes, which differ in structure and substrate requirements, it is possible to discriminate between their respective luminescent

reactions. Thus, in this particular system, the luminescence from the firefly luciferase reaction can be quenched while simultaneously activating the luminescent reaction of the Renilla luciferase. All experimental steps were carried out according to manufacturer's recommendation.

2.3. CELL CULTURE

Several different human and mouse cell lines were utilized in the various reporter gene assays. The cells were cultured in 5% CO^2 at 37°C using the following, optimal culture conditions [30]. NIH 3T3 (murine fibroblast) cells were grown in Dulbecos modified Eagles medium supplemented with 10% calf serum, 10 g/l streptomycin and 6 g/l penicillin as well as 2mM L-glutamine. HeLa cells (human cervical cancer), 661W (murine embryonic retinal) and aTN4 (murine lens) cells were cultured in Dulbecos modified Eagles medium supplemented with 10% fetal calf serum, 10 g/l streptomycin and 6 g/l penicillin as well as 2mM L-glutamine.

2.4. REVERSE TRANSCRIPTION PCR

Total RNA was purified from all three cell lines mentioned above using RNA-WIZZARD (Promega) and reverse transcribed using random hexamers and Superscript reverse transcriptase (Invitrogen). The oligonucleotides used for the amplification of Pax6, Actin and Gap3D transcripts were as follows; Pax6 (for all transcripts) forward (5'-TCCTTCACATCAGG-CATGTTGGGC-3'); reverse (5'-CCGGGAACTTGCACTG-GAAC-3'); Pax6 PO forward (5' CC-TCT-TTTCT-TATCGTTGAC-3'); Pax6 P1 forward (5'-GGAGTGAT-TAGTGGTTTGA-3'); Pa (5'-AGTTCATTCTCGTC-

TGGGTG-3'); Pax6 Exon 5 - (5'-GCTTGGTGGTGCTTT-GTCA-3'); Actin forward (5'ATGGT-GGGAATGGGTCAGAAGGAC-3'); reverse (5'-CTCTTTGATGTCAC-GCA-CGATTTC-3'); Gap3D forward (5'-AACGACCCCTTCATTGACCTC-3'); reverse (5'-ATCCACGACGGACACATTGG-3'). The cycling conditions were 94°C for 3 min, and then 30 cycles of 94°C for 30 sec, 55°C for 30 sec and 72°C for 1 min followed by a final step of 72°C for 5 min. PCR products were subsequently analyzed on 1% agarose gel (Fig. 7).

2.5. DNA TRANSFECTION

The respective cell lines were plated in a 96 well assay plate, at an optimal cell density, to achieve 40%-50% confluency the following day and were then transiently transfected using FuGENE 6 transfection reagent (Roche Molecular Biochemicals). The ratio between FuGENE reagent and DNA was set to 3:1 with 1µg of DNA used for every reaction according to manufacturer's recommendation. All constructs were prepared using commercial DNA preparation systems (DNA Mini-kit, Promega or EndoFree-Plasimid Maxi-kit) and DNA samples from two different DNA preparations were used for the transfection experiments. As an internal control, 100ng of pRL-TK (*Renilla reniformis* luciferase gene with HSV-TK promoter) were cotransfected. In order to account for experimental variation, five or more wells were transfected per individual construct in each independent experiment. In addition to the test fragments, a control vector carrying the SV-40 enhancer and promoter were included (Promega). Because, this combination of regulatory elements is supposed to drive reporter gene expression at high levels in all test cell lines, it was used an indicator for the success of the assay. To

assess transfection efficiency, a GFP reporter (pEGFP with CMV promoter, Clontech) was transfected and green fluorescence was visualized by fluorescent microscopy after 24h and 48h. The number of cells expressing the GFP protein at high levels usually ranged between 70%-80% of the total number of cells in each well. The cells were grown for 48h and then lysed in plates (1x lysis buffer, Promega). The assay was carried out using the Lmax Microplate Luminometer (Molecular devices) according to manufacturer's protocol and raw data was analyzed using standard statistical procedures.

2.6. IN SITU HYBRIDIZATION

In order to better understand the potential role of *ELP4* in aniridia we decided to use *in situ* hybridization to visualize its spatial and temporal expression pattern.

In situ hybridization has become a standard method used to analyze the developmental expression patterns of genes in a variety of organisms, including mouse. We chose to use dioxygenin-labeled single stranded RNA probes in our experiments because of convenience and because it did not involve the handling of radioactive isotopes. For all in situ experiments, we used the standard protocol developed by Wilkenson [31]. In order to obtain an accurate expression pattern and to identify background signal in the staining pattern, several probes were generated, directed to different parts of the *Elp4* transcript. In addition, to ensure good penetration of the probes into the tissue, full-length probes were hydrolyzed to the desired length according to standard procedures.

All *in situ* hybridization experiments were performed more than 3 times to ensure reproducibility of the results.

CHAPTER 3. RESULTS

3.1 CHARACTERIZATION OF ELP4

3.1.1. ELP4 EXPRESSION PATTERN

The data obtained from whole mount in situ hybridization experiments indicated a very broad, almost ubiquitous expression domain (Fig. 4). Using embryos from different stages of development, we observed this reproducible broad expression pattern, indicating that *Elp4* does not undergo an obvious developmental regulation. Furthermore, no apparent eye specific expression pattern was observed in our whole mount in situ hybridization experiments. This result is consistent with data obtained by Lauderdale et al. (unpublished) using northern blot analysis, which also showed *Elp4* expression in a wide variety of tissues but at greatly varying levels. After the *Elp4* expression domain was studied in whole-mount in situ hybridization the embryos were sectioned and analyzed for staining pattern. The sectioning confirmed the broad expression pattern but also revealed that the staining was not uniform. For example, as shown in figure 4, we observed an increased expression of *Elp4* in the dorsal neural tube as well as in the trigeminal ganglion (data not shown). Again, this confirmed results obtained previously by northern blot analysis.

Figure 4.

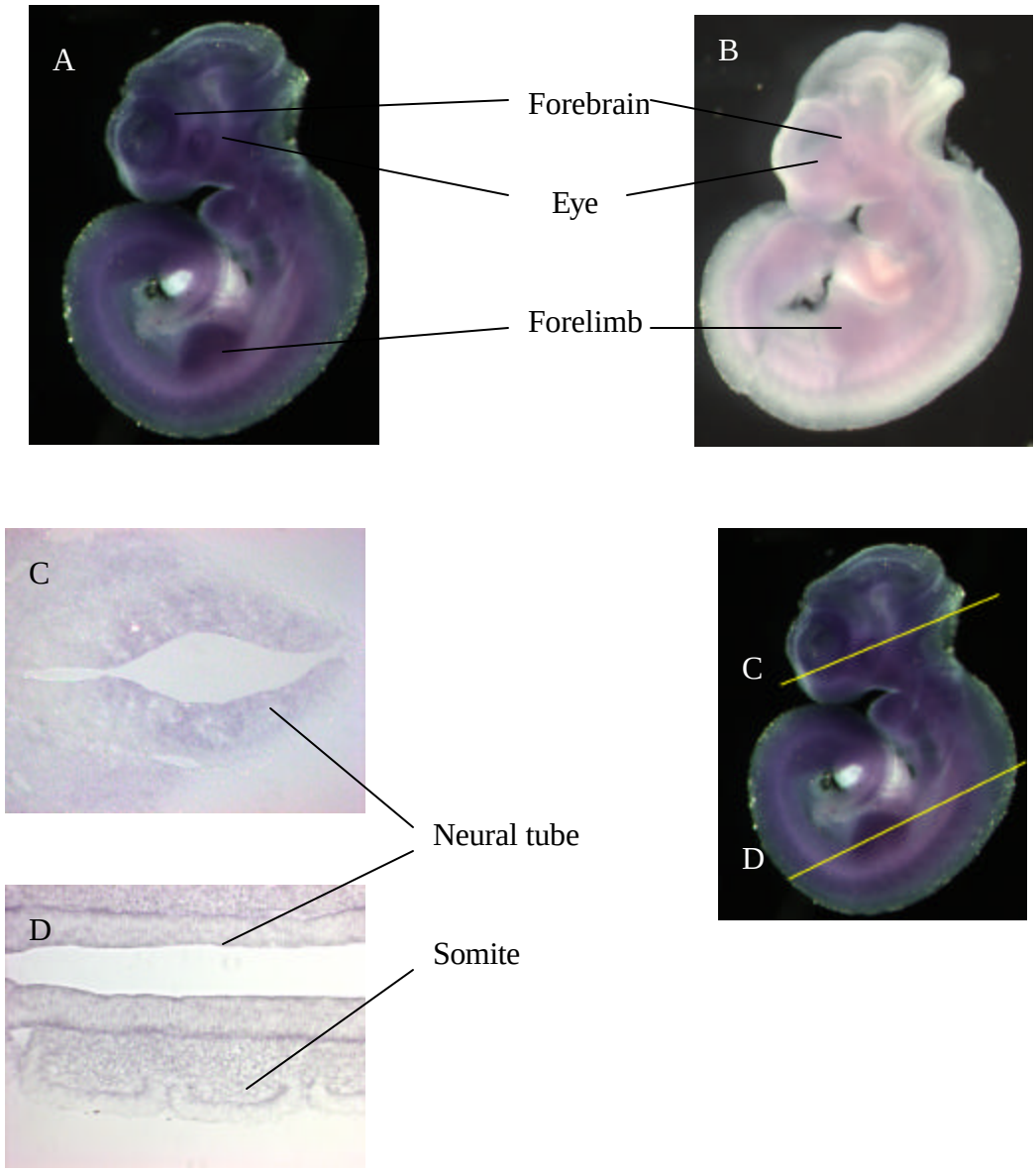


Figure 4. **Expression pattern of Elp4.** A-B Representative whole mount *in situ* hybridization of mouse embryo at E10.5 of development. A. Antisense probe for Elp4 mRNA shows broad, almost ubiquitious expression of Elp4. B. Sense probe for Elp4. C-D. Representive transverse cryo-sections (12µm) of mouse embryo at E10.5 of development after in situ hybridization for Elp4 mRNA. Sections show broad but not uniform expression pattern. High expression observed in dorsal neural tube compared to ventral neural tube. No expression is seen in the surface ectoderm. Plane of sections are indicated to the right. In situ hybridization was performed using ~300bp hydrolysed probe targeted to the 3' end of the transcript.

3.1.2. ELP4 PROMOTER ANALYSIS

Because of the location of the DRR in the intragenic region of *Elp4* and the associated question of how a Pax6 specific regulation is achieved, we analyzed the *Elp4* promoter and upstream region with regard to functional elements. Several studies in a variety of species, including mice and flies, demonstrated that the functional elements contained in the core promoter region of a given gene, determine the interaction with activator- or repressor proteins that bind to other regulatory sequences such as enhancers or repressors. As a result several models have evolved which describe the different mechanisms of how the core promoter structure can have an impact on this level of gene regulation [32]. Some of these models will be discussed below. It has become obvious in recent years that there is a distinct preference of certain enhancer elements to different kinds of promoters. We reasoned that knowing the functional components of the *Elp4* promoter, we would be able to propose and subsequently test a mechanistic model of how the DRR may exert its regulatory function in this genomic environment.

In order to determine the promoter region of *Elp4* we used existing data from 5' race experiments and compared different expressed sequence tags (EST) of *Elp4* for both mouse and human. A region containing part of exon 1 of the murine *Elp4* and an approximately 3.8 kb upstream region was cloned in both orientations into luciferase reporter vector pGL3 basic (Figure 5), which does not contain any regulatory sequences. In order to determine if other regulatory sequences were present in the 3.8 kb upstream region, deletion constructs of the original fragments were made containing varying sequences around the putative transcription initiation site. The data below represents the

Figure 5.

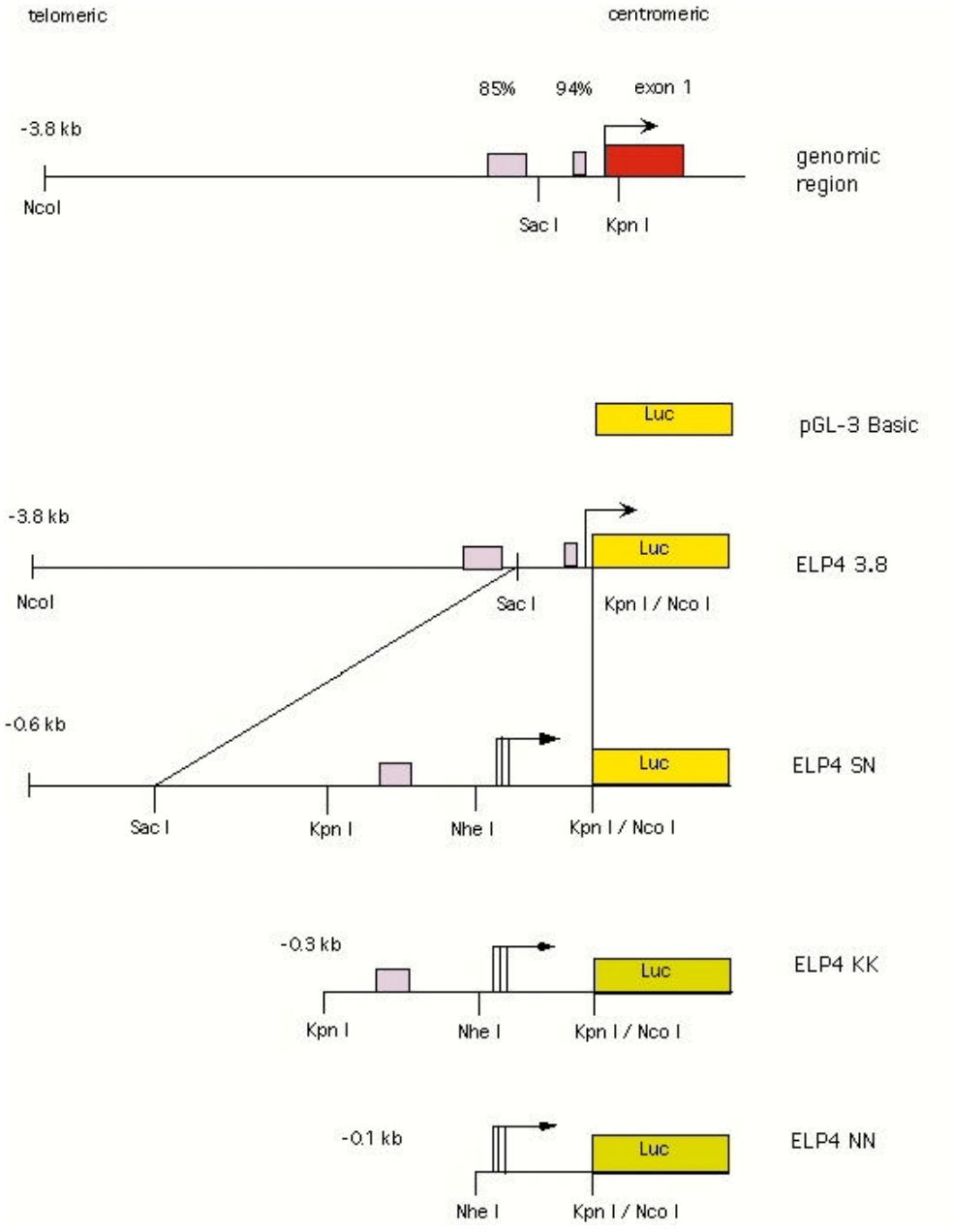


Figure 5**. Reporter constructs for Elp4 promoter analysis in transient transfection assay.** Various luciferase reporter constructs were generated to characterize the murine Elp4 promoter and promoter proximal region. A 3.8kb DNA fragment containing part of the murine Elp4 exon 1 and 3.8kb of upstream region was cloned into pGL-3 Basic reporter vector. Using the indicated restriction site a deletion series luciferase gene (Luc), yellow box; putative transcription start site, arrow; block evolutionary conserved sequence, shaded; fragment size indicated on left side; construct name indicated on right side. Elp4 exon 1 is shown as red box.

Figure 6.

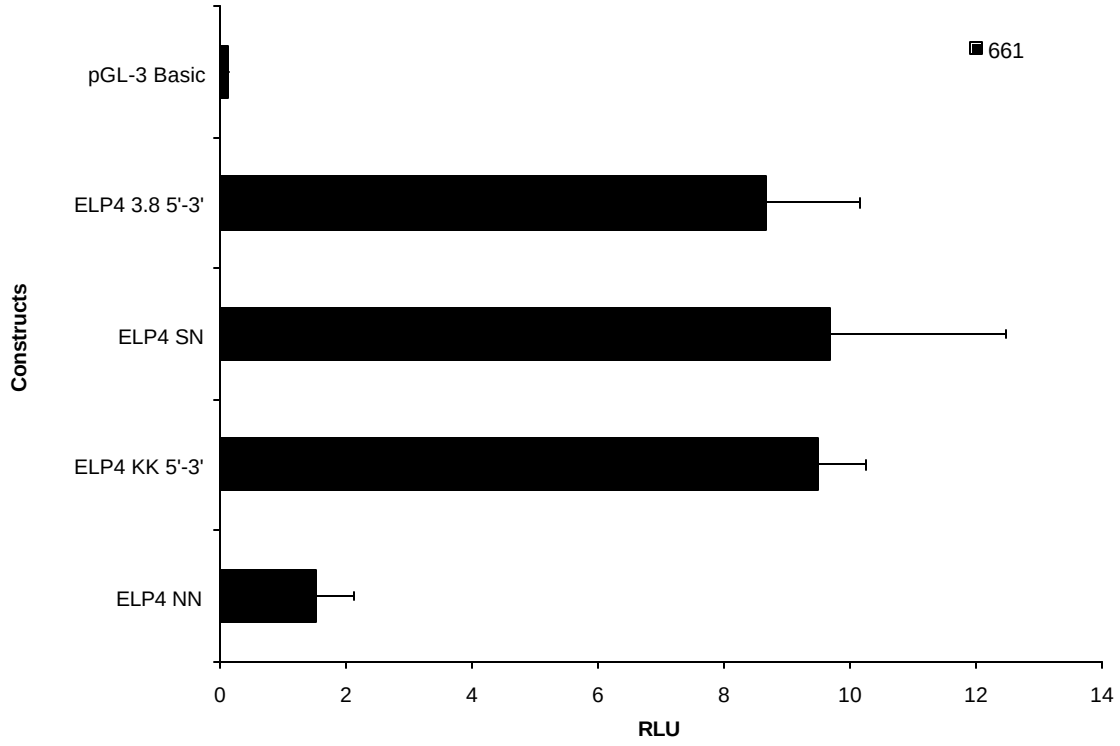
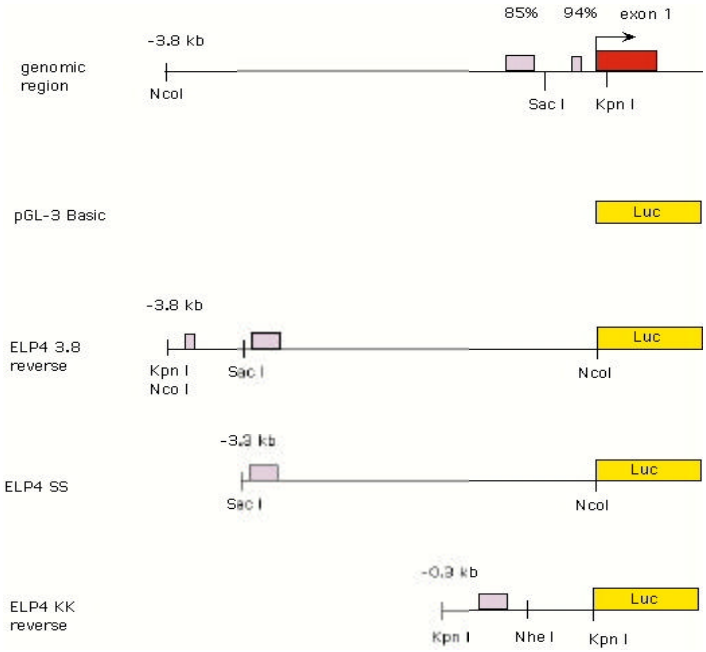


Figure 6**. Result of Elp4 promoter analysis in transient transfection assay**. The longest fragment (3.8kb upstream region) significantly increased reporter gene expression, compared to the pGL-3 Basic construct. 5' deletions of upstream region in ELP4 SN (550bp) and ELP4 KK (300bp) resulted in approximately the same reporter gene expression as the full length fragment, indicating that no other activating sequences are present in the upstream region. Further deletion down to 100bp (ELP4 NN) of upstream region resulted in significant decrease of fold activation compared to larger construct but still higher activation relative to pGL-3Basic construct. The data was obtained from three independent experiments with five internal repetitions using 661W cells (murine retinal). Average deviation of relative luminescence (RLU) is indicated by error bars.

Figure 7

A



B

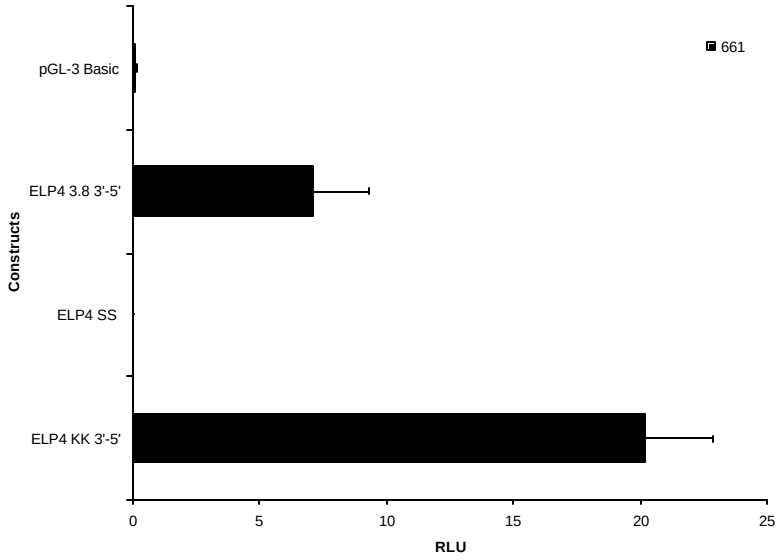


Figure. 7**. The Elp4 promoter acts bi-directionally**. Construct Elp4 3.8kb (3'-5') containing part of exon 1 of Elp4 and 3.8kb upstream region in the opposite orientation drove luciferase activity at high levels. Deletion of 550bp promoter region abolished activity (ELP4 SS). The Kpn I fragment (ELP4 KK 3'-5'), containing the evolutionary conserved sequence block, drove reporter gene expression at very high levels in the reverse orientation, confirming that the Elp4 promoter acts bi-directionally. Data shown was obtained from three individual experiments with five internal repetitions. Cell lines used were 661W (murine retinal). Average deviation of relative luminescence (RLU) is indicated.

values of at least three independent experiments with five internal repetitions. When the fragment containing part of exon 1 of *Elp4* and 3.8 kb (*ELP4* 3.8) of upstream region was transiently transfected into 661W (murine retinal cell line), we observed a marked increase of firefly luciferase activity compared to the pGL-3 Basic construct, which served as the baseline due to the lack of any regulatory sequences (Fig. 5, 6). Similar results were obtained in two independent experiments using HeLa cells (human cervical cancer) (data not shown). A 550 bp Sac I-NcoI fragment containing the putative promoter and promoter proximal region was found to drive the reporter gene at similar levels as the full-length fragment (Fig. 6). Further deletion down to a 300 bp Kpn I fragment (*ELP4* KK) containing a sequence box of 94% identity between human and mice, did not result in a significant drop in reporter gene activation (Fig. 6). In contrast to that, deletion of additional 200bp from this fragment including this highly conserved box resulted in a significant decrease in activation, but reporter gene expression was still significantly higher than that of the pGL-3 Basic construct (Fig. 6). Therefore, this 100bp fragment (*ELP4* NN) is likely to be the minimal *Elp4* promoter.

As a control, we included the 3.8 kb full-length fragment in the reverse orientation and expected to observe no luciferase activity over the pGL-3 Basic construct levels. This reasoning was based on the assumption that the putative promoter region of *Elp4* was now located 3.8 kb upstream of the luciferase gene and that in the reverse orientation relative to the native genomic arrangement.

Surprisingly, when this construct was tested in our assay we observed very high levels of luciferase activity indicating that the promoter functions independent of its orientation (Fig 7B). Because our previous experiments indicated that, the 550 bp Kpn I fragment

contained all promoter and promoter proximal elements we decided to delete this sequence from the 3.8 kb full-length fragment in the reverse orientation. We expected that if most of the activity were residing in this region then deleting it would demonstrate the necessity of this region along with verifying results obtained with the reverse construct. As expected, luciferase activity dropped to pGL-3 Basic levels after deletion of the SacI-NcoI fragment from the reverse fragment demonstrating that this region is required for gene activation in our assay (7B). The fragment to conferring very high activity had been shown to be the 300 bp KpnI fragment (*ELP4* KK) containing the highly evolutionarily conserved sequence block. When tested in the opposite orientation relative to the genomic arrangement, we again observed high levels of luciferase activity confirming our previous results obtained with the full-length fragment (7B).
These results indicate that the *Elp4* promoter can act in an orientation independent fashion and that the core promoter sequence is contained within the 100 bp KpnI fragment (*ELP4* NN).

3.1.3. SEQUENCE ANALYSIS OF *ELP4* PROMOTER

After defining the functional *Elp4* promoter region, we searched for the usual sequence motifs associated with general promoter structures such as TATA or CAAT boxes. Using the software program MatInspector [33] we compared the human and mouse *Elp4* promoter and promoter proximal regions against each other. Using phylogenetic footprinting (see below), we were able to identify two blocks of highly conserved sequence. One short box of 40bp (Fig. 8) with a sequence identity of 94% between mouse

Figure 8.

Figure 8. **Sequence analysis of Elp4 promoter and upstream region.** A. Evolutionarily conserved mammalian C-type LTR CCAAT box. The degree of conservation is 94% over 40bp between human and mouse. Deletion of this box results in significant reduction of reporter gene expression in our assay indicating a functional significance. B. Block of conserved sequence (85% human/mouse) approximately 1kb upstream of putative transcription start site. Locations of potential transcription factor binding sites are indicated. C. Core promoter sequence of Elp4 determined by deletion analysis. Core promoter sequence contains two potential Inr and respective DPE consensus sequences.

and human was identified just upstream of the putative transcription start site. When analyzed with MatInspector, the sequence revealed a conserved mammalian C-type LTR CAAT box. As mentioned above, deletion of the region containing this motif caused the luciferase activity of the reporter construct to drop significantly, indicating a functional role of this motif. The second block of conserved sequence was located approximately 800 bp upstream of the putative transcription start site in both mouse and human. Several potential regulator-binding sites were identified between mouse and human, having the same relative position with respect to the genomic organization in both species. Binding sites for the following factors were found; Nix 2.4, Octane, Pour-factor and Msx-1, 2 as well as TALE-class homeodomain binding sites were found in this region. Interestingly, no consensus TATA-box sequence was found in the region indicating that the *Elp4* promoter belongs to the class of TATA-less promoter, which are usually associated with housekeeping genes. Furthermore, we were able to identify an evolutionarily conserved binding site for the Core Promoter-Binding Protein (CPBP) within the defined 100bp minimal promoter region (*ELP4* NN). This transcription factor has been found to bind specifically to the core promoter sequence of TATA-less genes.

3.1.4. CONCLUSION

The broad expression pattern seen in the whole-mount in situ hybridization and the absence of a specific staining pattern in the eye suggested that *Elp4* is probably not involved in the human panocular disease aniridia. Recent studies in yeast characterized the *Elp4* (elongator protein 4) homolog as part of the elongator complex, which is associated with the hyperphosphorylated C-terminal domain of polymerase II during

transcription elongation. Our data from in situ hybridization experiments and promoter analysis supports the idea that *Elp4* is in fact a housekeeping gene.

3.2. IDENTIFICATION OF ENHANCER ELEMENTS IN THE DOWNSTREAM REGULATORY REGION OF PAX6

3.2.1. PHYLOGENETIC FOOTPRINTING

Evolutionary sequence comparison (phylogenetic footprinting) has been proven to be a very useful tool in the identification of new regulatory sequences for a variety of genes [34]. According to the hypothesis, non-coding regions, displaying an unusually high sequence homology between different species, may have a regulatory function and were therefore preserved throughout evolution. Furthermore, factors binding to these regions are predicted to be similar as well. Although phylogenetic footprinting analysis does not allow a sufficiently high resolution to pin point individual binding motifs for transcription factors, it can predict distinct regulatory sequences, which span a number of potential binding sites. Using cross-species comparison of sequences from mice and humans has the advantage of a relative close evolutionary distance of about 60 million years with most of the transcription factor binding sites arranged in conserved blocks of similar sequence. The degree of conservation can be considerable, reaching up to 90% for specific genes. Nevertheless, because of the small evolutionary distance among mammals and the relatively slow rate of divergence in this class, more non-coding sequences may retain sequence identity then is actually functional which will in all likelihood complicated the identification of relevant regulatory sequences. In our search for

upstream regulators of Pax6, previous data obtained from aniridic patients and YAC transgenesis experiments (see above) prompted us to focus our attention on a genomic, non-coding region approximately 150 kb downstream of the Pax6 transcription unit. A phylogenetic footprinting analysis of this region between human and mouse revealed a number of highly conserved sequence blocks of up to 96% identity (Fig. 9). For our analysis, a sequence is considered highly conserved if it is more than 100bp long and has more than 75% identity. Although, we initially used only human/mouse sequence comparison, the availability of new sequence data from a variety of species allowed us to extend our search for conserved regulatory region to those species as well. As shown in figure 9 most of the conserved blocks identified between human and mouse are also highly conserved between several other species.

3.2.2. CHARACTERIZATION OF EXPERIMENTAL CELL LINES

Because, we were looking for Pax6 specific regulatory sequences, we chose two Pax6 expressing cell lines and one non-expressing cell line. 661W, a murine embryonic retinal cell line and alpha TN4 a cell line derived from the lens of transgenic mice were chosen to test our putative enhancer fragments. As a negative control, we used NIH 3T3 cells, a murine fibroblast line that does not express Pax6.

We first sought to characterize these cell lines with respect to Pax6 expression and the presence of alternative transcripts since Pax6 is being transcribed from three different promoters. As expected, the Pax6 gene is only expressed in 661W and aTN4 cells and not in NIH 3T3 cells (Fig. 10). Furthermore, RT-PCR also shows that of the three Pax6

Figure 9.

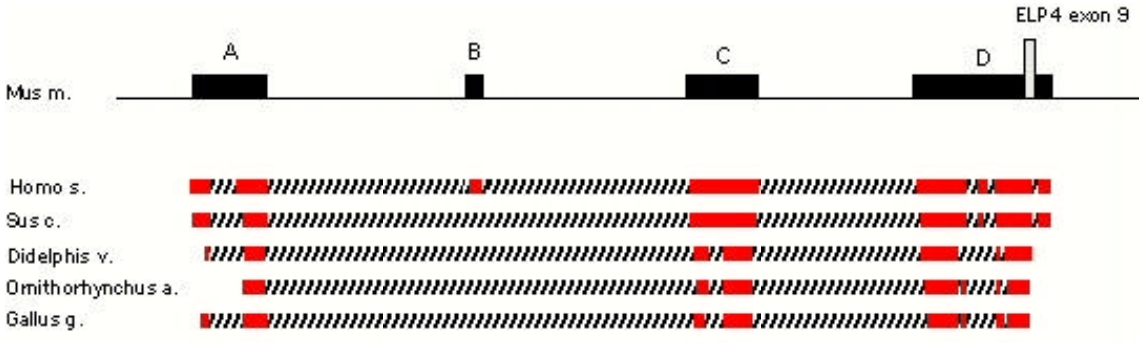


Figure 9**. Identification of evolutionarily conserved sequences in the DRR.** Alignment of genomic sequences from several species including human, pig, opossum, platypus and chicken reveals blocks of highly conserved sequences in the mouse DRR. Mouse DRR region shown on top with conserved sequences shown as black boxes and Elp4 exon shown as grey boxes. In the multiple-species alignment, solid red areas represents sequences of 90%-100% conservation whereas dashed line indicates a degree of conservation below 10%. *Homo s*., human; *Sus c*., pig; *Didelphis v*., opossum; *Ornithorhynchus a*., platypus; *Gallus g*., chicken

Figure 10.

A

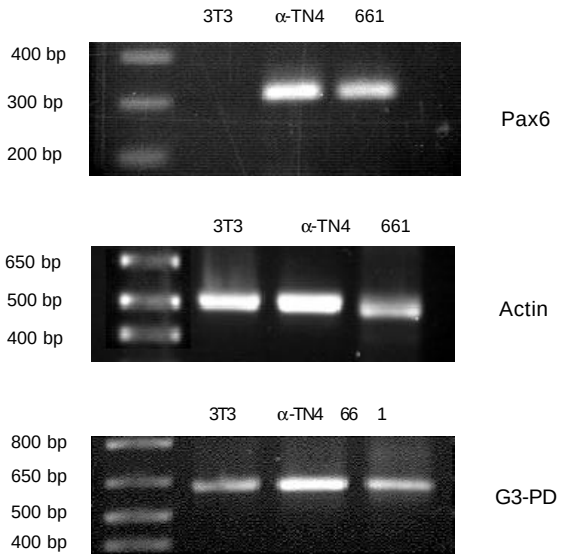
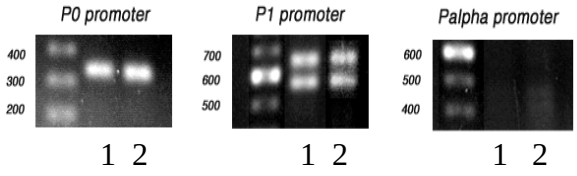


Figure 10. Characterization of experimental cell lines. A. RT-PCR using specific primers for Pax6, Actin and G3-PD. Pax6 is only expressed in aTN4 and 661 cells but not in 3T3 cells. Both Actin and G3-PD, known to be house-keeping genes, are expressed in all three cell lines.
B. RT-PCR using primers specific for the three alternative transcription start sites. Lane 1 indicates 661 cell line and lane 2 indicates aTN4 cell line. PO and P1 (two bands observed due to alternative splicing) promoter are utilized in both Pax6 expressing cell lines whereas the P -alpha does not seem to be utilized at all.

B



promoters only PO and P1 were utilized; Pa does not seem to be used for transcript initiation.

3.2.3. ENHANCER ANALYSIS

Several of the conserved sequence blocks (Fig. 9), found through evolutionary sequence comparison were tested in our dual luciferase reporter assay. All data represented below was obtained from at least three independent experiments in 661W (murine retinal). In the course of this study, it became apparent that our primary promoter, the SV40 promoter, drove luciferase gene expression at extremely high levels in the aTN4 lens cells, probably due to the presence of an activating factor in this cell line. Therefore, we decided to include Hela cells, a human cancer cell line, which expresses Pax6, in our experiments. When tested in our cell culture system, element B did not enhance expression of the luciferase gene over the promoter level, indicating that either no regulatory sequence is present in this region or our test cell lines do not provide the correct molecular environment for this element to be active (Fig. 11B). Furthermore, a 5.7kb fragment containing element D showed enhancement at significant but low levels (Fig. 11B). Element C on the other hand clearly enhanced reporter gene expression in our assay with both the SV40 and the TK promoter (Fig. 12 and data not shown). In order to identify the minimal sequence required for increased reporter gene expression a deletion series of element C was prepared (Kim et al. personal communication) The minimal fragment contained approximately 300bp of highly conserved sequence in which the activator binding site probably resides. When tested in 661 retinal cells this minimal

Figure 11.

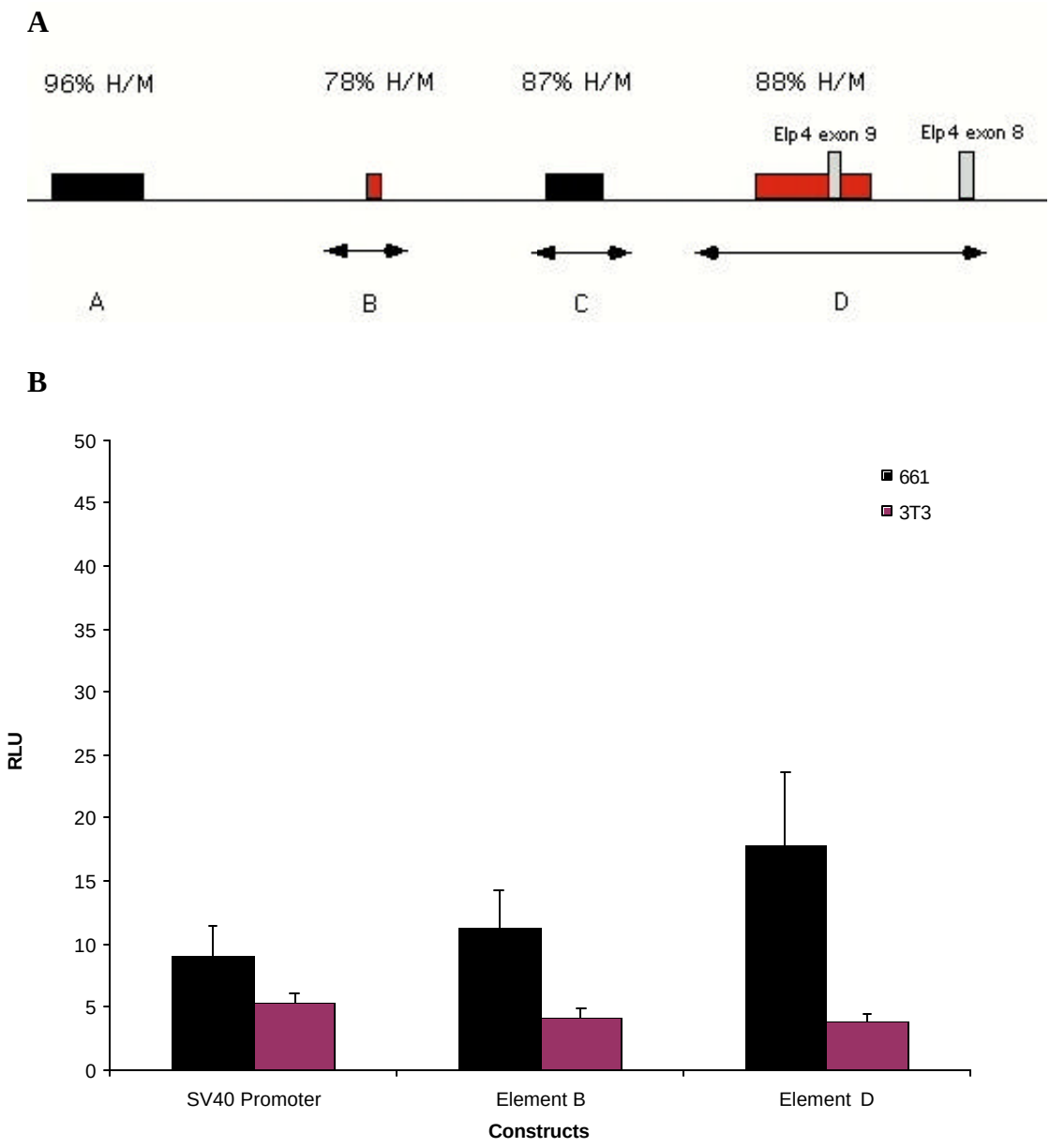


Figure 11. **Transient transfection analysis of putative regulatory elements B and D.** A. Putative regulatory regions, identified by phylogenetic footprinting, are shown as red or black boxes. Red boxes indicate elements tested in the luciferase assay shown in panel B. Constructs were prepared by inserting element B and D into pGL-3 SV40 promoter vector. Elp4 exons are shown as grey boxes. Conservation is shown above the corresponding boxes. Black arrows indicate the region amplified by PCR and tested in reporter assay. B. Results of reporter assay for element B and D. Reporter activity given in relative luminescence units (RLU). Data shown was obtained from three to five individual experiments with five internal repetitions. Cell lines used were 661W (murine retinal) and NIH 3T3 (murine fibroblast). Average deviation shown as arrow bars.

Figure 12.

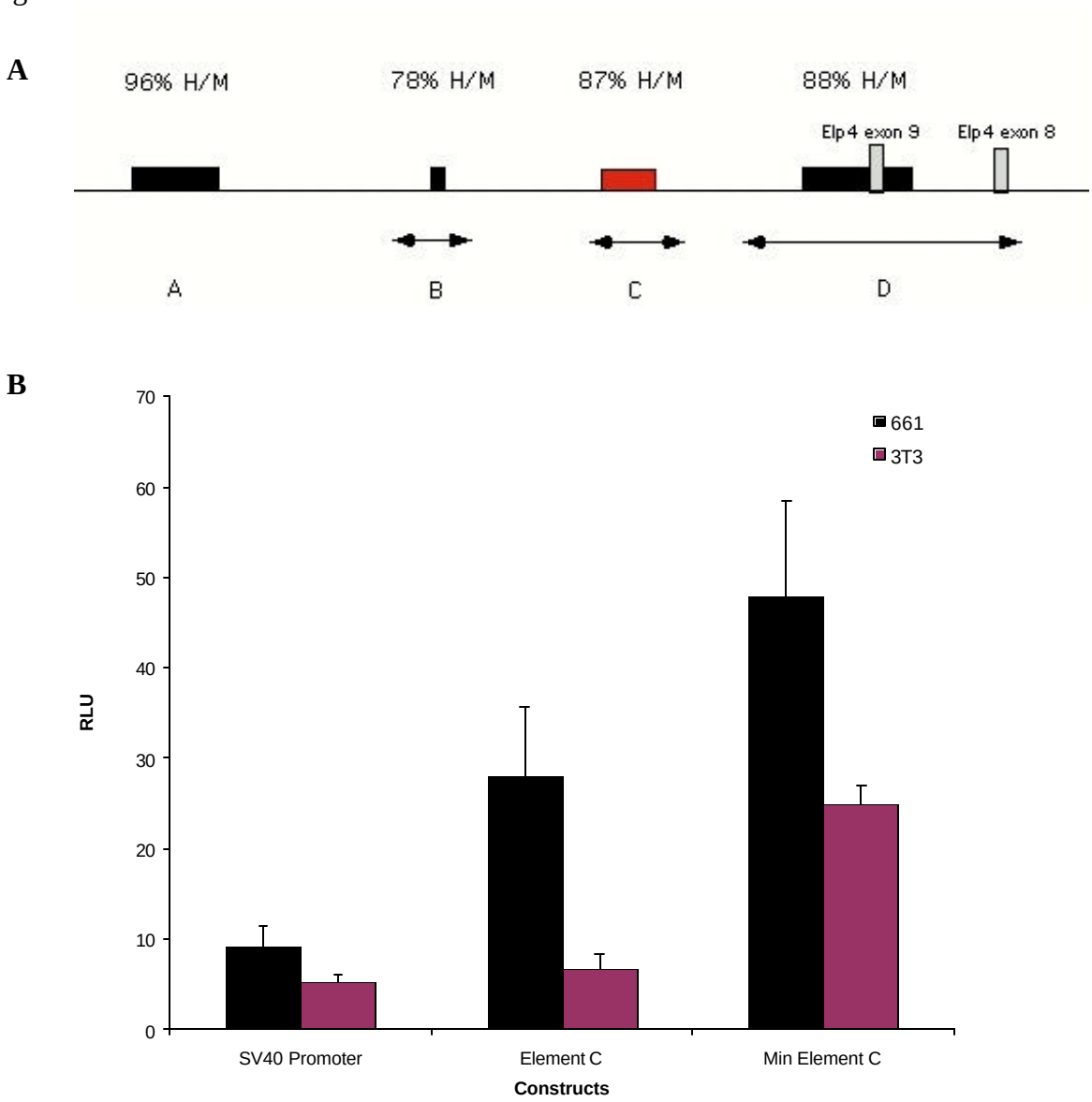


Figure 12. **Transient transfection analysis of putative regulatory element C.** A. Putative regulatory regions identified by phylogenetic footprinting are shown as red or black boxes. Red boxes indicate elements tested in the luciferase assay shown in panel B. Constructs were prepared by inserting element C into pGL-3 SV40 promoter vector. Min Element C comprises 600bp of the 5' region of element C and was obtained by deletion of the 3' sequence of element C. Elp4 exons are shown as grey boxes. Degree of conservation is shown above the corresponding boxes. Black arrows indicate the region amplified by PCR and tested in reporter assay. B. Results of reporter assay for element C derivative. Reporter activity given in relative luminescence units (RLU). Data shown was obtained from three to five individual experiments with five internal repetitions. Cell lines used were 661W (murine retinal) and NIH 3T3 (murine fibroblast). Average deviation shown as arrow bars.

fragment drove reporter gene expression at even higher levels than the original fragment indicating that a repressive sequence was removed as a result of the sequential, directed truncation of element C (Fig. 12). Surprisingly, a similar result was obtained when the minimal element was tested in NIH 3T3 cells, a cell type which does not express Pax6 as shown by RT-PCR (Fig. 12). It is possible that the repressive element deleted in the bashing experiment is actually responsible for repressing the activator function in cells where Pax6 is usually not expressed.

3.2.4 SEQUENCE ANALYSIS OF ELEMENT C

Sequence analysis of element C revealed the presence of several potential transcription factor-binding sites, which were evolutionarily conserved and therefore may have a functional role in gene regulation (Fig. 13). Among the candidate binding proteins were the Pbx-1/Meis-1 heterodimer complex and the Cone-Rod-Homeobox transcription factor, both being crucial to eye development. All Pbx-1/Meis-1 binding sites were present in mouse, human and chick whereas only one of the CRX binding sites was seen in all three species (Fig. 13). Especially interesting was the observation that one Pbx-1/Meis-1 binding site was situated within the minimal element C. The fact that this minimal sequence drove reporter gene expression at high levels in both Pax6 expressing and non-expressing cells may be accounted for by the presence of a repressor binding site in the full length element C, which usually prevents activity of this enhancer in certain tissues. Consistent with this interpretation is the finding that both Meis-1 and Pbx-1 are expressed in NIH 3T3 fibroblast cells, thereby strengthening the argument that this factor complex is responsible for the observed reporter gene activity in our assay [35].

Figure 13.

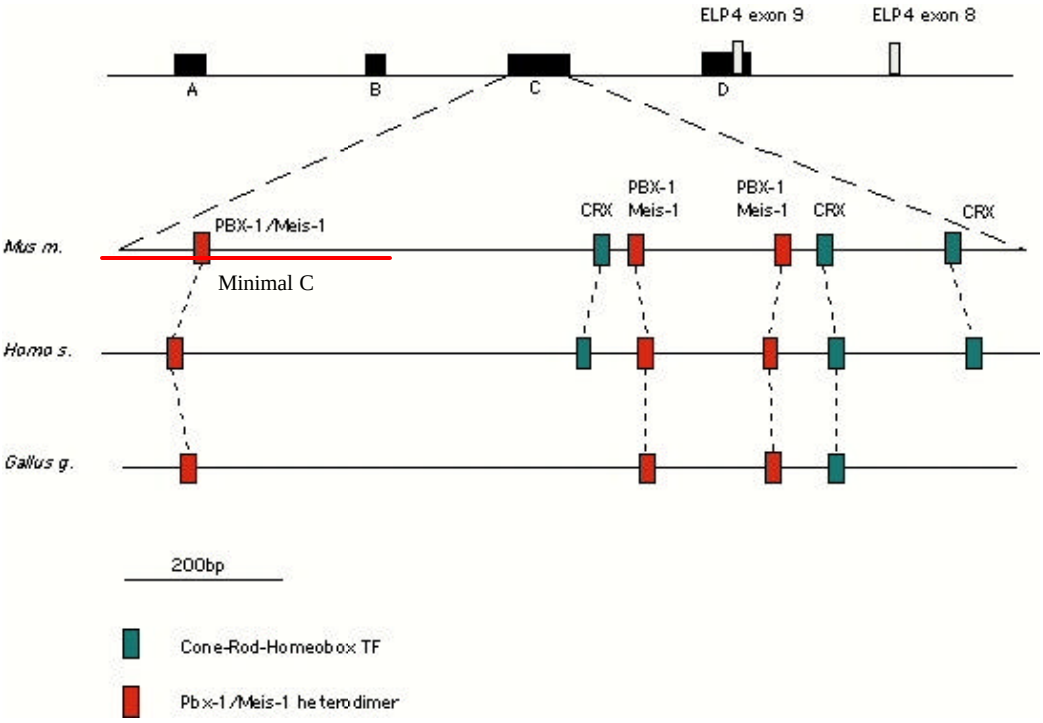


Figure 13. **Sequence analysis of enhancer element C**. Representation of mouse DRR is on top with regions of high evolutionary conservation displayed as black boxes. Elp4 exons are shown as grey boxes. Sequence analysis using MatInspector software revealed the presence of several potential regulator binding sites (core similarity 1.0) in element C such as Pbx1/Meis1 heterodimer and the Cone-Rod transcription factor as indicated. *Mus musculus*, house mouse; *Homo sapiens*, human; *Gallus gallus*, chicken.

CHAPTER 4. DISCUSSION

In this study, we attempted to further the understanding of the recently described downstream regulatory region (DRR) of the Pax6 transcription factor. Not only has this regulatory region been shown to be crucial to Pax6 expression, but it is also very interesting as regards the regulatory mechanisms at work.

Because the DRR is located at a great distance from the putative target and also because it is located within the intron of another gene, understanding how this region exerts its function in this genomic context would not only further our understanding of the way Pax6 is regulated but also shed light on more general gene regulatory principles. In order to address this very complex question, we chose to pursue two complementing approaches. As discussed above, the presence of additional 3' regulatory sequences for Pax6 was first deduced from patients suffering from the panocular disease aniridia.

A subset of these patients not only lacked those crucial regulatory elements, but also showed heterozygous loss of a second gene downstream of the Pax6 transcription unit. It became apparent that before a reasonable model for the function of the DRR can be proposed, this second gene would have to be characterized thoroughly. Therefore, we decided to first assess the spatial and temporal expression pattern of the novel gene, now termed *Elp4* (Elongator protein 4). Whole-mount in situ hybridization on mouse embryos of varying developmental stages using dioxigenin-labeled RNA probes directed against the *Elp4* mRNA showed a very broad expression domain (Fig. 4A). Sections obtained from these embryos confirmed the broad expression pattern, and revealed locations of

higher *Elp4* expression. For example, higher levels of *Elp4* transcript were observed in parts of the developing nervous system including the dorsal neural tube (Fig. 4C) and the trigeminal ganglion. These findings are consistent with data previously obtained from northern blot analysis, which showed the presence of *Elp4* transcript in a wide variety of tissues but at different levels. The absence of a specific eye expression pattern and the fact that *Elp4* is so widely expressed made it unlikely that *Elp4* is actually involved in any of the aniridia cases mentioned above. In fact, recent studies in yeast suggest that the homolog of *Elp4* in this species functions as part of the elongator complex, associating with RNA polymerase II during transcription elongation [36-38]. This finding is consistent with expression data we observed, as a protein involved in a process like transcription would probably be expressed ubiquitously.

Because the functional elements contained in the promoter and immediate upstream region of a given gene can provide information about how the gene may be regulated we decided to characterize the *Elp4* promoter. This issue is particularly interesting, considering that, Pax6 and *Elp4* promoters are located approximately equidistant from the DRR, but it was supposed that only Pax6 is controlled by regulators residing in this region.

The promoter analysis revealed two highly conserved sequence blocks upstream of the putative transcription start site, containing the same transcription factor binding site in both mouse and human (Fig. 5). Curiously, when the *Elp4* promoter and immediate upstream region was tested in our luciferase reporter assay, we observed an orientation independence, which could be assigned to a 300bp fragment just upstream of the putative transcription start site (Fig. 6A, 7A,). Inclusion of additional upstream sequence,

especially the second block of conserved sequence (200bp, 85% human/mouse), did not enhance reporter gene expression. This may indicate that the functionally relevant regulatory elements are contained in the region surrounding the immediate transcription start site. Alternatively, it is possible that these regulator binding sites are only active in specific tissues, and that our test cell lines may not provide the correct molecular environment for them to act. This interpretation would also fit our findings because additional regulatory elements would account for our results that increased *Elp4* transcript levels can be observed in certain locations in the developing mouse embryo by in situ hybridization such as the dorsal neural tube (Fig. 4C).

Interestingly, sequence analysis of the proximal block of conserved sequence (40bp, 94% human/mouse) revealed the presence of evolutionarily conserved mammalian type-C LTR CAAT box (Fig. 8A). Our inspection did not reveal a consensus TATA sequence prompting us to believe that *Elp4* transcription is initiated from an initiator (INR) sequence. Indeed, sequence analysis of the 100bp minimal fragment (*Elp4* NN), required to drive reporter gene expression, and showed two possible Inr consensus sequences in close proximity to each other (7C). Inspection of the region +28 to +32 from the putative Inr(s) also revealed two potential DPE (Downstream promoter element) consensus sequences. Comparison of this region with cDNA from the mouse EST database confirmed the initiation of several transcripts within a reasonable nucleotide window relative to the Inr sequences described above. We therefore propose that the *Elp4* core promoter consists of at least one Inr sequence, which is probably coupled with a DPE. In addition to these elements, a proximal CAAT box may enhance transcription from the *Elp4* promoter.

Although CAAT boxes are not symmetrical as regards their sequence, they have been shown to act as strong transcription enhancing elements in an orientation independent fashion; this would explain the results observed in our experiments. If the *Elp4* promoter is in fact, a bidirectional promoter one would expect the presence of another gene in the proximity of the defined *Elp4* promoter transcribed in the opposite orientation relative to *Elp4*. To address this question, we screened the NCBI (National Center for Biotechnology Information) EST database using the genomic region of *Elp4*, including exon 1 and 3.8kb of upstream sequence. As expected we observed a second transcript in the predicted orientation located just upstream of the defined *Elp4* promoter. The identity or functional significance of this second transcript is not known so far, but taken together these observations support the hypothesis that the *Elp4* promoter is in fact a TATA-less bidirectional promoter. The concept of bidirectional gene organization is still somewhat obscure, and much remains unclear about the regulatory mechanisms at work. After the completion of the human genome project, it has become more and more apparent that a divergent (bidirectional) gene configuration is a common theme in the human genome. Up to date, there are twenty two described loci where genes have been shown to be driven through a bidirectional promoter region e.g. BRCA1/NBR2, DNA-PKcs/MCM4 [39]. The functional significance of those arrangements remains unclear but one possibility is that it may allow genes to utilize the same regulatory elements for the purpose of co-regulated expression. In fact, many of the examples of bi-directional promoters so far are found to be associated with DNA repair genes [40], strengthening the argument for a possible co-regulation. This concept is somewhat reminiscent of the bacterial gene regulation through the organization of functionally linked genes in operons

[41]. In parallel to the analysis of the *Elp4* gene, we also tried to further dissect the functional elements contained in the DRR. Several enhancer sequences have been identified already using mainly reporter gene experiments in transgenic mice as well as DNase I hypersensitivity assays. All regulatory elements described so far in this region are associated with evolutionarily conserved sequences (human/mouse). The sequence conservation over such a huge evolutionary distance (60 million years) is unlikely to be coincidental, since non-functional sequence is not subject to selective pressure and would diverge over time. Therefore, it is very likely that conserved regions are of functional significance and probably contain binding sites for regulatory proteins such as transcriptional activators and repressors. It is also reasonable to assume that the interacting protein factors are either homologous or at least related between the compared species, indicating the conservation of a given developmental pathway or mechanism. Using phylogenetic foot printing, we were able to identify four blocks of highly conserved sequence in the region previously found to be crucial to Pax6 expression. Of the three elements tested only element C displayed pronounced enhancer activity in a Pax6 positive cell line used in our assay. No increased reporter gene expression was observed with this full-length fragment, in Pax6 negative 3T3 fibroblast cells; this data is consistent with the hypothesis that this enhancer indeed acts on the Pax6 transcription unit specifically. Interestingly, the fact that a minimal enhancer fragment derived from the full-length element C could drive the luciferase reporter gene at high levels in non-Pax6-expression fibroblast cells indicates that binding sites for both positive and negative regulators are present in this region. Using the sequence analysis tool MatInspector we were able to predict several potential binding factors located within the minimal enhancer

element C. Among those were Meis I and Pbx I; both have been shown to be upstream regulators of Pax6 in the eye [1, 42]. Consistent with our results is a finding by Kleinjan et al. in which a considerably larger fragment from this region, containing element C derived from human genomic DNA, could drive ß-galactosidase reporter gene expression in the neural and pigmented retina of transgenic mice [26].

The fact that of our three test elements only one was found to drive reporter gene expression at enhanced levels clearly shows a shortcoming of this approach. Several possibilities may account for our results. First, Pax6 is known to be expressed in a variety of locations other than the eye. Because we used mainly cell lines derived from murine ocular tissue, our cell lines may not provide the required molecular environment for the putative enhancer element to function properly. A second possibility is that not all conserved sequences may be functionally significant, although evolutionary sequence comparison has now been well established as a tool to identify novel regulatory sequences. A third possibility would be that for a given enhancer element to function, it has to be located in a broader genomic context to allow interaction with other regulatory elements.

Having obtained additional information regarding the DRR we are now in the position to postulate and test possible mechanism of how the DRR exerts its regulatory function in the region. It has become apparent that some enhancers clearly have a core-promoter specificity or at least a preference [43, 44]. Extensive studies in a variety of species has resulted in the identification of several motifs within the core promoter sequence, the most common of which are the TATA box, initiator (Inr) and downstream core promoter element (DPE). While each of these elements has a specific function in the transcription

process it is also true that not all promoters contain the same functional elements. In other words," there is no universal core promoter". Furthermore, different elements can be arranged differently to achieve modular transcription control at this level. The emerging idea behind this core promoter variety seems to be the creation of a completely new level of gene regulation. For example, genes may have acquired alternative core promoter elements in order to avoid repression or activation by certain regulatory elements such as enhancers or repressors [43, 44]. This may be particularly desirable for genes located in clusters.

In theory, this same rationale can be applied to the DRR because of the genomic environment in which it is located. The arrangement of Pax6 on the centromeric and *Elp4* on the telomeric side is evolutionarily conserved down to fish but how is it possible that the DRR is only activating the Pax6 transcription unit and not the one of *Elp4* [27]? One possibility is that as discussed above, the enhancer elements located in the DRR display a core-promoter requirement and can only interact with a core promoter sequence similar to that observed with the endogenous Pax6 promoters. Alternatively, the DRR elements may prefer the Pax6 promoters but are capable of interacting with the *Elp4* promoter as well. A third possibility which has to be considered is that a boundary element is present between the DRR and *Elp4* promoter region, shielding it from transcriptional activation by the DRR [45]. Nevertheless, no evidence in support of this model has been reported so far.

Considering the facts, we favor the first model in which the elements contained in the DRR can only interact with the endogenous Pax6 promoters but not with the *Elp4* promoter due to requirement of certain core-promoter elements. In fact, when comparing

the regulatory motifs of PO, P1 and Pa with those of the *Elp4* promoter the differences are striking [46]. In the Pax6 PO and P1 promoter a CCAAT box and a TATTAA box consensus sequence can be observed with no Inr consensus sequence been present (Figure 14). On the other hand, the *Elp4* promoter is obviously a TATA-less promoter which contains an consensus Inr sequence and a DPE as well as a conserved CCAAT box in the promoter proximal region. Taken together, these facts argue strongly for possible promoter specificity rather than a boundary element in order to restrict DRR action to the Pax6 transcription unit.

Figure 14.

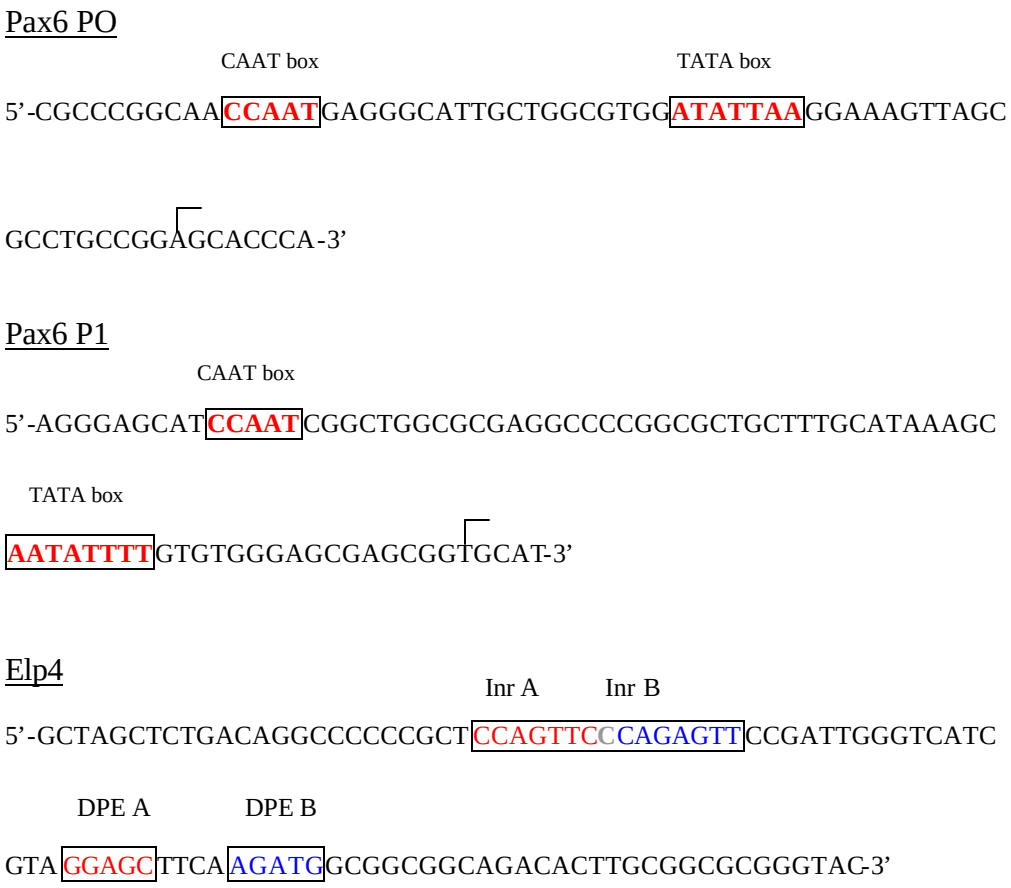


Figure 14. **Comparison of regulatory elements within two endogenous Pax6 promoters and the Elp4 promoter.** Position of regulatory elements as indicated in the various promoters. Initiator (Inr) and corresponding downstream promoter element (DPE) were detected in the Elp4 promoter whereas Pax6 promoters contained both CAAT and TATA boxes. The Pa-promoter of Pax6 was excluded for the lack of bonafied data. Arrows indicate transcription initiation site.

REFERENCES

1. Jean, D., et al., *Molecular regulators involved in vertebrate eye development.* Mechanisms of development., 1998. **76(1-2)**: p. 3-18.
2. Wawersik, S., et al., *Vertebrate eye development as modeled in Drosophila.* Human molecular genetics., 2000. **9(6)**: p. 917-25.
3. Ton, C.C., et al., *Small eye (Sey): cloning and characterization of the murine homolog of the human aniridia gene.* Genomics., 1992. **13(2)**: p. 251-6.
4. Gruss, P., C. Walther, and D.o.M.C.B.G.G. Max Planck Institute for Biophysical Chemistry, *Pax in development.* Cell., 1992. **69(5)**: p. 719-22.
5. Walther, C., P. Gruss, and D.o.M.C.B.G.G. Max Planck Institute for Biophysical Chemistry, *Pax-6, a murine paired box gene, is expressed in the developing CNS.* Development (Cambridge, England), 1991. **113(4)**: p. 1435-49.
6. Glaser, T., et al., *Genomic structure, evolutionary conservation and aniridia mutations in the human PAX6 gene.* Nature genetics., 1992. **2(3)**: p. 232-9.
7. Kozmik, Z., et al., *Alternatively spliced insertions in the paired domain restrict the DNA sequence specificity of Pax6 and Pax8.* The EMBO journal., 1997. **16(22)**: p. 6793-803.
8. Singh, S., et al., *Iris hypoplasia in mice that lack the alternatively spliced Pax6(5a) isoform.* Proceedings of the National Academy of Sciences of the United States of America., 2002. **99(10)**: p. 6812-5.

9. Marquardt, T., et al., *Pax6 is required for the multipotent state of retinal progenitor cells.* Cell., 2001. **105(1)**: p. 43-55.

10. Simpson, T.I., et al., *Pax6; a pleiotropic player in development.* BioEssays : news and reviews in molecular, cellular and developmental biology., 2002. **24(11)**: p. 1041-51.

11. Hill, R.E., et al., *Mouse small eye results from mutations in a paired-like homeobox-containing gene.* Nature., 1991. **354(6354)**: p. 522-5.

12. Sander, M., et al., *Genetic analysis reveals that PAX6 is required for normal transcription of pancreatic hormone genes and islet development.* Genes & development., 1997. **11(13)**: p. 1662-73.

13. Kioussi, C., et al., *Pax6 is essential for establishing ventral-dorsal cell boundaries in pituitary gland development.* Proceedings of the National Academy of Sciences of the United States of America., 1999. **96(25)**: p. 14378-82.

14. Stoykova, A., et al., *Forebrain patterning defects in Small eye mutant mice.* Development (Cambridge, England), 1996. **122(11)**: p. 3453-65.

15. Mastick, G.S., et al., *Pax-6 functions in boundary formation and axon guidance in the embryonic mouse forebrain.* Development (Cambridge, England), 1997. **124(10)**: p. 1985-97.

16. Jones, L., et al., *Pax6 is required for the development of the forebrain axonal connections.* Development (Cambridge, England), 2002. **129(21)**: p. 5041-52.

17. Stoykova, A., et al., *Pax6-dependent regulation of adhesive patterning, R-cadherin expression and boundary formation in developing forebrain.* Development (Cambridge, England), 1997. **124(19)**: p. 3765-77.

18. Anderson, T.R., et al., *Differential Pax6 promoter activity and transcript expression during forebrain development.* Mechanisms of development., 2002. **114(1-2)**: p. 171-5.

19. Kammandel, B., et al., *Distinct cis-essential modules direct the time-space pattern of the Pax6 gene activity.* Developmental biology., 1999. **205(1)**: p. 79-97.

20. Xu, P.X., et al., *Regulation of Pax6 expression is conserved between mice and flies.* Development (Cambridge, England), 1999. **126(2)**: p. 383-95.

21. Fantes, J., et al., *Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype.* Human molecular genetics., 1995. **4(3)**: p. 415-22.

22. Hanson, I.M., et al., *PAX6 mutations in aniridia.* Human molecular genetics., 1993. **2(7)**: p. 915-20.

23. Hanson, I., et al., *Missense mutations in the most ancient residues of the PAX6 paired domain underlie a spectrum of human congenital eye malformations.* Human molecular genetics., 1999. **8(2)**: p. 165-72.

24. Schedl, A., et al., *Influence of PAX6 gene dosage on development: overexpression causes severe eye abnormalities.* Cell., 1996. **86(1)**: p. 71-82.

25. Lauderdale, J.D., et al., *3' deletions cause aniridia by preventing PAX6 gene expression.* Proceedings of the National Academy of Sciences of the United States of America., 2000. **97(25)**: p. 13755-9.

26. Kleinjan, D.A., et al., *Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6.* Human molecular genetics., 2001. **10(19)**: p. 2049-59.

27. Kleinjan, D.A., et al., *Characterization of a novel gene adjacent to PAX6, revealing synteny conservation with functional significance.* Mammalian genome : official journal of the International Mammalian Genome Society., 2002. **13(2)**: p. 102-7.

28. Sambrook, *Molecular Cloning, A Laboratory Manual.* Third Edition ed. Third Edition. Vol. 1. 2001, Cold Spring Habour: Cold Spring Habour Laboratory Press.

29. Marchuk, D., et al., *Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products.* Nucleic acids research., 1991. **19(5)**: p. 1154.

30. Lindl, T., *Zell-und Gewebekultur*. 4 ed. Vol. 1. 2000, Heidelberg, Berlin: Spektrum Akademischer Verlag. 260.

31. and S.-W.a. Moser, *In situ hybridization.* Histochemistry, 1993. **100**(100): p. 431-400.

32. Smale, S.T., M. Department of, and U.o.C.L.A.S.o.M.U.S.A.s.h.u.e. Immunology, *Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes.* Biochimica et biophysica acta., 1997. **1351(1-2)**: p. 73-88.

33. Quandt, K., et al., *MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data.* Nucleic acids research., 1995. **23(23)**: p. 4878-84.

34. Mèuller, F., et al., *Search for enhancers: teleost models in comparative genomic and transgenic analysis of cis regulatory elements.* BioEssays : news and reviews in molecular, cellular and developmental biology., 2002. **24(6)**: p. 564-72.

35. Chang, C.P., et al., *Meis proteins are major in vivo DNA binding partners for wild-type but not chimeric Pbx proteins.* Molecular and cellular biology., 1997. **17(10)**: p. 5679-87.

36. Kim, J.H., et al., *Human Elongator facilitates RNA polymerase II transcription through chromatin.* Proceedings of the National Academy of Sciences of the United States of America., 2002. **99(3)**: p. 1241-6.

37. Winkler, G.S., et al., *Elongator is a histone H3 and H4 acetyltransferase important for normal histone acetylation levels in vivo.* Proceedings of the National Academy of Sciences of the United States of America., 2002. **99(6)**: p. 3517-22.

38. Winkler, G.S., et al., *RNA polymerase II elongator holoenzyme is composed of two discrete subcomplexes.* The Journal of biological chemistry., 2001. **276(35)**: p. 32743-9.

39. Adachi, N. and M.R. Lieber, *Bidirectional gene organization: a common architectural feature of the human genome.* Cell., 2002. **109(7)**: p. 807-9.

40. Connelly, M.A., et al., *The promoters for human DNA-PKcs (PRKDC) and MCM4: divergently transcribed genes located at chromosome 8 band q11.* Genomics., 1998. **47(1)**: p. 71-83.

41. McClure, W.R., *Mechanism and control of transcription initiation in prokaryotes.* Annual review of biochemistry., 1985. **54**.

42. Zhang, X., et al., *Meis homeoproteins directly regulate Pax6 during vertebrate lens morphogenesis.* Genes & development., 2002. **16(16)**: p. 2097-107.

43. Aso, T., et al., *A TATA sequence-dependent transcriptional repressor activity associated with mammalian transcription factor IIA.* The EMBO journal., 1994. **13(2)**: p. 435-45.

44. Mack, D.H., et al., *Specific repression of TATA-mediated but not initiator-mediated transcription by wild-type p53.* Nature., 1993. **363(6426)**: p. 281-3.

45. Labrador, M., V.G. Corces, and T.J.H.U.N.C.S.B.M.D.U.S.A. Department of Bio