uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Nicholas Shipley

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Biology)

GRADE / DEGREE

Department of Biology

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Role of the I56i Enhancer Sequence in the Expression of the *Dlx* Gene Family

TITRE DE LA THÈSE / TITLE OF THESIS

Dr. M. Ekker

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Dr. D. Johnson

Dr. M. Jonz

Dr. B. Willmore

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Role of the I56i Enhancer Sequence in the Expression of the *Dlx* Gene Family

Nicholas Shipley

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
University of Ottawa
in partial fulfillment of the requirements for the
M.Sc. degree in the

Ottawa-Carleton Institute of Biology

Thèse soumise à
Faculté des études supérieures et postdoctorales
Université d'Ottawa
en vue de l'obtention de la maîtrise ès sciences

L'Institut de biologie d'Ottawa-Carleton

Your file Votre référence
ISBN: 978-0-494-41684-6
Our file Notre référence
ISBN: 978-0-494-41684-6

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Marc Ekker for providing me an opportunity to work with him in his lab. He provided me with excellent supervision, support, and of course funding and space for me to complete this research.

I would like to thank Dr. Luc Poitras for support and teaching throughout my entire project. His helpful and patient explanations were invaluable and his advice on all subjects was appreciated immensely.

Thanks to Gary Hatch, for his constant willingness to answer any question and provide any help needed. Thanks to Ryan MacDonald for support, advice, and guidance, both in the lab and in the outside world. I would like to thank the rest of the members of the Ekker lab for their discussions, advice, help and friendship.

I would like to thank Dr. Marie-Andrée Akimenko for providing space to do mouse work as well as Adrianna Gambarotta and Sylvie Emond for all their help with the mice. Thanks to Dr. Doug Johnson and Iain Lambert for being on my committee.

Finally, I would like to thank my family and friends for their support, advice, motivation, and frequent patience.

ABSTRACT

The genes of the *distal-less* (*Dlx*) gene family encode homeodomain transcription factors involved in the development of the brain, branchial arches, limbs, teeth, olfactory placodes and otic vesicles. In mice and humans, six *Dlx* genes have been identified and are organized in bigene clusters. The I56i *cis*-acting regulatory element is located within the intergenic region of the *Dlx5/Dlx6* locus. This element has been shown to be active in the forebrain (telencephalon and diencephalon), the first and second branchial arches, and the apical ectodermal ridge. In this study, we identified eight putative transcription factor binding regions within the sequence of the I56i enhancer using *DNaseI* footprinting. These included putative binding sites for GATA-1, Engrailed-1, DLX and homeodomain transcription factors. Three to eight base-pair mutations were introduced into these putative binding sites. The effect of mutagenesis of four of these putative binding sites was evaluated *in vivo* in transgenic mice using a *lacZ-β*-globin minimal promoter construct and *in vitro* using electrophoretic mobility shift assays (EMSA). Activity of the mutant enhancer in the forebrain of mouse embryos was reduced in some of the transgenic animals obtained. A reduction or complete loss of activity was seen in the branchial arches of all but one of the transgenic mice carrying the mutant enhancer constructs. EMSA demonstrated that proteins from the nuclear extracts of E11.5 mouse brains and E13.5 mouse forebrains bind fragments of DNA corresponding to the I56i enhancer sequences. Further, mutagenesis of the putative binding sites appears to reduce the ability of proteins to bind these areas of the I56i enhancer sequences. These experiments provide preliminary evidence suggesting that a number of factors may be involved in regulating the activity of the I56i enhancer and demonstrating the

potential role of the I56i enhancer in the regulation of gene expression in the forebrain and branchial arches.

RÉSUMÉ

Les gènes de la famille *distal-less* (*Dlx*) qui codent pour des facteurs de transcription à boîte homéo sont impliqués dans le développement du cerveau, des arcs branchiaux, des membres, des dents, des placodes olfactives et des vésicules otiques. Chez la souris et l'humain, six gènes *Dlx* ont été identifiés. Ils sont organisés en tandem de gènes. L'élément régulateur agissant en *cis* (enhancer), I56i, est situé à l'intérieur de la région intergénique du locus *Dlx5/Dlx6*. Il a été montré que cet élément était actif dans le cerveau antérieur (télencéphale et diencéphale), le premier et second arc branchial de même que dans l'arête apical ectodermale. Dans cette étude, nous avons identifié huit régions potentielles pouvant lier des facteurs de transcription à l'intérieur de l'élément I56i par l'analyse d'empreinte à la *DNase I*. Ces régions de liaison incluent des sites potentiels pour les facteurs de transcription GATA-1, Engrailed-1, DLX et pour des facteurs à boîte homeo. Des mutations de trois à huit paires de bases ont été introduites dans ces sites de liaison potentiels. Les effets de la mutagénèse de quatre sites de liaison ont été évalués *in vivo* par la production de souris transgénique utilisant une construction contenant le promoteur minimal de la β–globine de même que le gène rapporteur *lacZ*. De plus, l'effet de la mutagénèse a été évalué *in vitro* par des essais de retard de mobilité sur gel (EMSA). L'activité de l'enhancer mutant dans le cerveau antérieur d'embryons de souris était réduite dans quelques animaux transgéniques obtenus. Chez la majorité des souris transgéniques à l'exception d'une seule, possédant les constructions sous le contrôle des enhancers mutants, une réduction ou encore une perte complète a été observée dans les arcs branchiaux. Les essais de retard sur gel ont démontré que certaines protéines, provenant d'un extrait nucléaire de cerveau de souris de temps de gestation E11.5 et E13.5, pouvaient lier des fragments d'ADN correspondant à la séquence

de l'élément I56i. De plus, la mutagénèse des sites de liaison potentiels semble réduire l'habilité de ces protéines à lier ces régions de l'élément I56i. Les résultats préliminaires de ces expériences suggèrent qu'un bon nombre de facteurs semblent être impliqué dans la régulation de l'activité de l'élément I56i et démontrent le rôle potentiel de l'élément I56i dans la régulation de l'expression génique dans le cerveau antérieur et les arcs branchiaux.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ABBREVIATIONS AND ACRONYMS

AER: apical ectodermal ridge
ASD: autistic spectrum disorder
BMP: bone morphogenetic protein
bp: base pairs
BT: basal telencephalon
Di: diencephalon
DNA: deoxyribonucleic acid
dATP: deoxyadenosine triphosphate
ddH_2O: double distilled H_2O
dNTP: deoxynucleotide triphosphate
E10.5: embryonic development, 10.5 days past conception
E11.5: embryonic development, 11.5 days past conception
E12.5: embryonic development, 12.5 days past conception
E13.5: embryonic development, 13.5 days past conception
EDTA: ethylene diamine tetraacetic acid
EMSA: electrophoretic mobility shift assay
FGF: fibroblast growth factor
FN: frontonasal prominence
FP: Protected site in *DNase*I footprint
GABA: γ-amino butyric acid
H: human
HDTF: homeodomain transcription factor
HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hy: hyoid arch
III: third ventricle of diencephalon
IPTG: Isopropyl β-D-1-thiogalactopyranoside
Kb: kilobases
L: AER of limbs
LB: luria broth
LGE: lateral ganglionic eminence
LJ: lower jaw
LV: lateral ventricle
M: mouse
Md: mandibular arch
MGE: medial ganglionic eminence
mRNA: messenger RNA
Mut: site containing mutation corresponding to FP site
MZ: mantle zone
NCX: cortex
NP-40: Nonidet P-40
PBS: phosphate buffered saline
PCR: polymerase chain reaction
PCX: palliocortex

pdIdC: Polydeoxy(Inosinate-Cytidylate) Acid
RNA: ribonucleic acid
S: *Spheroides nephelus*
SDS: sodium dodecyl sulfate
SNP: single nucleotide polymorphism
St: somites
SVZ: subventricular zone
T: *Takifugu rubripes*
TAE: Tris-acetate EDTA
TBE: Tris-borate EDTA
TE: Tris-EDTA
tRNA: transfer RNA
UJ: upper jaw
vbf: vibrissae
VT: ventral telencephalon
VZ: ventricular zone
Z: *Danio rerio* (zebrafish)

1. INTRODUCTION

1.1 Transcription Factors

Transcription factors are proteins required in the initiation of transcription in eukaryotes. These proteins are responsible for a number of transcription-related functions, including the assembly of the RNA polymerase on the DNA and aiding in the dissociation of the DNA strands. Transcription factors may also bind DNA upstream of the transcription initiation site, such as at the promoter, and interact with the transcription complex to regulate transcription. Homeobox genes encode one class of transcription factors, homeodomain-containing proteins. The homeodomain is a conserved DNA binding domain of approximately 60 amino acids. Numerous homeobox genes have been identified and these genes play a major role during development (Alberts 2002).

The *distal-less* (*Dll*) and *distal-less* related genes (*Dlx*) comprise one family of homeodomain transcription factors. The *distal-less* gene, identified in *Drosophila* in the 1980s, encodes a homeodomain transcription factor that plays a key role in the development of the fly (Cohen, Bronner et al. 1989). Genes related to the *Dll* gene have been identified in a number of other invertebrates as well as many vertebrates. The *Dlx* genes also encode a homeodomain transcription factor and are similarly required in development.

1.2 Regulatory Elements

Although transcription is initiated by transcription factors, the regulation of the rates of transcription often occurs through other mechanisms, including via regulatory elements, or *cis*-acting regulatory elements (Griffiths 2000; Alberts 2002). Regulatory elements are non-coding stretches of DNA that are involved in the regulation of transcription. These elements help control the level of gene expression as well as the spatial and temporal location of

expression. Although the mode of action of the regulatory elements is not known, it has been proposed that proteins, or *trans*-acting factors, binding to these elements interact with the proteins bound to the promoter, resulting in a change in transcription. Regulatory elements may exert their effect on a gene many thousands of base pairs from the element. It is believed the DNA forms a loop in such a manner that the proteins bound to the regulatory element are positioned close enough to exert their influence on the proteins bound to the promoter.

In general, two classes of *cis*-acting regulatory elements have been identified. Enhancers are regulatory elements that activate or increase the rate of transcription. Conversely, silencers reduce the rate of transcription. The presence of multiple regulatory elements affecting a single gene allows for a wide range of potential signals, and combinations of signals, to be produced by a single gene. By varying the way in which a gene is expressed in different cells, regulatory elements aid in the development of different signalling pathways without the need for an increase in the number of coding genes.

1.3 Conservation of Regulatory Elements - Ultraconserved Sequences

Given the role of regulatory elements in the regulation of transcription, it has been suggested that these elements must be maintained within the genome in order to maintain the regulatory pathways in which they are involved. In order to ensure they remain functional, it is believed these elements are subjected to negative selection, limiting variations to these sequences (Waterston, Lindblad-Toh et al. 2002; Katzman, Kern et al. 2007). Bejerano et al. identified 481 segments of DNA that were 100% conserved when comparing the human, rat and mouse genomes (Bejerano, Pheasant et al. 2004). Of these, more than half are in regions for which evidence of transcription (expressed sequence tags or mRNAs) have not been identified and more than 150 are found in the intergenic region between genes. A much

larger proportion of these genes are linked to developmental functions than would be expected under random circumstances. Although not all ultraconserved regions have been fully characterized, many have been identified as *cis*-acting regulatory elements (Pennacchio, Ahituv et al. 2006).

To test the theory that regulatory elements remain ultraconserved due to their fundamental role in development, Ahituv et al. removed four separate ultraconserved enhancers from the mouse genome (Ahituv, Zhu et al. 2007). The enhancers selected were located close to genes that showed a clear phenotype when inactivated or altered. However, no abnormalities were identified in any of the mice produced lacking the ultraconserved elements. The authors suggest that a phenotype may become evident following a longer timescale or it may be masked by redundancy in regulatory elements. Nevertheless, this suggests that the strong negative selective pressure that has maintained the sequence of these elements is not related to their regulatory role. However, other studies have shown that small changes in the sequence of a regulatory element can effect the activity of this element (discussed below in section 1.10). This variability in the effect of changes to regulatory elements demonstrates the difficulty in analyzing the role of regulatory elements in development and the potentially complex regulatory pathways in which they may be involved.

1.4 The *Distal-less* (*Dll*) Gene

The *distal-less* gene is required for the correct development of the distal appendages in *Drosophila*, including the limbs and antennae. In fact, *Distal-less* is one of the earliest molecular markers in the developing limb primordia (Cohen 1990). During embryogenesis, *Dll* transcripts are seen in cells of the developing limbs and antennae. Expression of *Dll* is

also seen in other appendages, such as the mouthparts and analia, sensory structures, and the central nervous system, including parts of the brain and ventral nerve cord (Panganiban 2000). The gene spans 20kb of DNA and contains seven exons. The 61 amino acid homeodomain spans across two of the seven exons. These two exons are split by an intron (Vachon, Cohen et al. 1992).

Drosophila mutants lacking functional copies of the *Dll* gene die during embryogenesis. This may be due to the lack of development of larval limbs or another undetected disruption in the development of the central nervous system (Cohen and Jurgens 1989). Furthermore, mutations that reduce the levels of *Dll* in developing *Drosophila* larvae result in a range of defects in the limbs, dependent on the relative levels of *Dll* still expressed. These defects result in the fusion and reduction in size of limb structures. This suggests that a graded amount of *Dll* is required to ensure the proper patterning of the limbs (Cohen, Bronner et al. 1989).

Dll mutants show defects in other areas of the fly, showing a role for this gene outside of the distal appendages. This includes sensory structures, such as the antennal, maxillary, labial and labral sense organs of the larval mouth, and Keilin's organ (Cohen and Jurgens 1989). Ectopic expression of *Dll* can also induce ectopic leg or antenna formation, dependent on the location of the expression and presence or absence of other factors, such as the Homothorax (Hth) transcription factors (Panganiban 2000).

In *Drosophila*, various factors are involved in the regulation of *Dll*. During embryogenesis, the Wingless (Wg) protein is required for *Dll* activation in the thoracic primordium. It is also partially responsible for expression in the embryonic head, along with Engrailed (En). Other proteins required for the correct expression of *Dll* in the embryo

include Decapentaplegic (Dpp), and the Hox proteins Ultrabithorax (Ubx), Abdominal A (AbdA) and Deformed (Dfd). In the larva, both Wg and Dpp are required for the maintenance of *Dll* expression in the larval leg and antennal imaginal disc. The direct targets of Dll are unknown, but many transcription factors lie genetically downstream of *Dll*. It is also believed that Dll forms complexes with other transcription factors in order to interact with and affect the expression of target genes (Panganiban 2000).

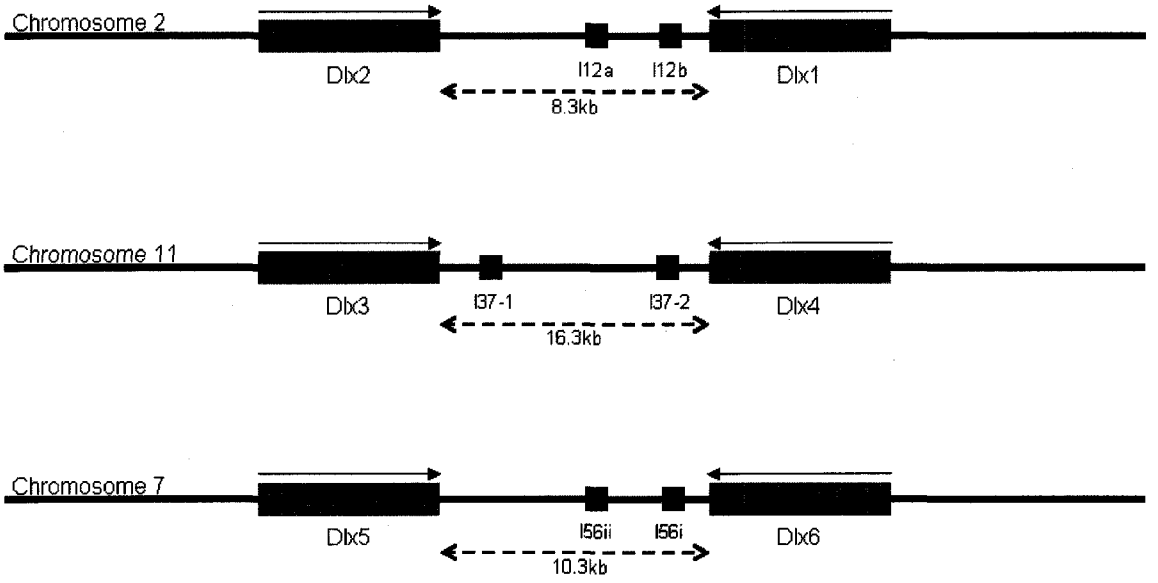
1.5 Vertebrate *Dlx* Genes

A wide number of lineages outside of arthropods have homologues to the *distal-less* gene. While invertebrates have only one *distal-less* gene, up to eight *distal-less* related genes, *Dlx*, have been identified in vertebrates. The vertebrate *Dlx* genes share a strong similarity in their 61 amino acid homeodomain when compared to the *Drosophila* ortholog, *Dll*. The consensus sequences for all *Dlx* genes show 78% similarity when compared to the *Drosophila Dll* gene. Other areas of sequence similarity exist between vertebrate *Dlx* genes and the *Drosophila Dll* gene, but these are relatively short areas (Panganiban and Rubenstein 2002). Furthermore, a conserved DNA binding site (5'-(A/C/G)TAATT(G/A)(C/G)-3') has been identified that is similar among many of the *Dlx* proteins (Feledy, Morasso et al. 1999).

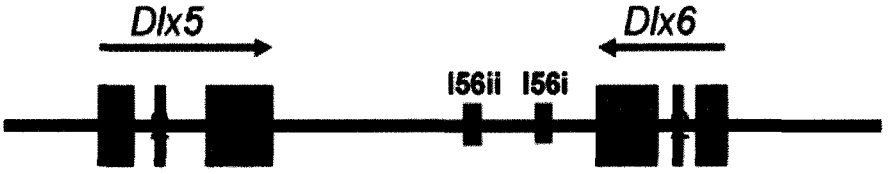
The conservation of the vertebrate *Dlx* genes extends beyond sequence similarity to the genomic organization within the vertebrates. The vertebrate *Dlx* genes are generally organized in three convergently transcribed bigene clusters: *Dlx1/Dlx2*, *Dlx3/Dlx4*, and *Dlx5/Dlx6* (Figure 1.1) and are found on distinct chromosomes. In mice and humans, these clusters are found on chromosomes 2, 11 and 7, respectively (Stock, Ellies et al. 1996). The intergenic region between the two genes of a cluster is relatively short, ranging from 3kb to

Figure 1.1. A) Organization of the vertebrate *Dlx* bi-gene clusters. Genes are shown in blue and the direction of transcription is indicated by an arrow above the genes. One gene family (*Dlx2*, *Dlx3*, *Dlx5*) is shown on the left and the other family (*Dlx1*, *Dlx4*, *Dlx6*) is on the right. Intergenic distances, as seen in mouse, are shown by the dashed arrow below. Relative positions of the *cis*-acting regulatory elements found in the intergenic regions are shown in red. The URE2 element, found 12kb upstream of *Dlx1* is not shown. **B) Detailed view of *Dlx5/Dlx6* locus, including organization of the exons.** The pattern of three exons is seen in the other two *Dlx* bigene clusters (not shown).

A)


Chromosome 2
Dlx2
I12a
I12b
Dlx1
8.3kb
Chromosome 11
Dlx3
I37-1
I37-2
Dlx4
16.3kb
Chromosome 7
Dlx5
I56ii
I56i
Dlx6
10.3kb

B)


Dlx5
Dlx6
I56ii
I56i

15kb in human, mouse, zebrafish, and two species of pufferfish (Ghanem, Jarinova et al. 2003). Zebrafish present a small exception to this pattern in that, in addition to the six genes mentioned above, they also contain two additional *Dlx* genes. Unlike the other *Dlx* genes in vertebrates, these additional *Dlx* genes in zebrafish are not arranged in a bigene cluster with each other or with other known *Dlx* genes. Some of the more primitive vertebrates also show exceptions to this rule of three bigene clusters. Only a single *Dlx* gene has been identified in *Amphioxus* and three *Dlx* genes have been found in tunicates (Panganiban and Rubenstein 2002).

The six common *Dlx* genes are also organized into two sub-families based on the amino acid similarities within their homeodomains. One sub-family consists of *Dlx1*, *Dlx4* and *Dlx 6* and the other sub-family of *Dlx2*, *Dlx3*, and *Dlx5* (Stock, Ellies et al. 1996). The bigene clusters described in the previous section are each composed of one member from each of the *Dlx* sub-families. The commonalities among the *Dlx* genes have led to the suggestion that all clusters originated from a single cluster or a single gene ancestor (Stock, Ellies et al. 1996). Also, the patterning of *Dlx* gene organization has been used as a tool to predict their evolutionary origin.

1.6 Evolution of the *Dlx* Genes

Homeodomain conservation, similar genomic organization, and the variation among the number of *Dlx* genes found in different species have led to the theory that the six *Dlx* genes found in vertebrates have evolved via a series of genomic duplications. This is supported by the linkage of *Dlx* gene clusters to *Hox* gene clusters. Three *Dlx* gene clusters found in higher vertebrates are each linked to a unique *Hox* cluster (Stock, Ellies et al. 1996; Ellies, Stock et al. 1997; Amores, Force et al. 1998; Force, Lynch et al. 1999). This genomic

organization is also seen in the invertebrate *Caenorhabditis elegans*, in which its single *Dlx* gene is linked to a single *Hox* cluster. It is believed that a tandem duplication event resulted in the configuration seen in the tunicate *Ciona intestinalis*, which contains a convergently transcribed *Dlx* gene pair linked to a single *Hox* cluster (Stock, Ellies et al. 1996; Sumiyama, Irvine et al. 2003). Subsequent genomic duplications resulted in multiple sets of *Dlx* gene pairs being linked to *Hox* clusters, such as is found in vertebrates. Research has shown a number of variations to this pattern of genomic organization, as seen in *Drosophila*, *C. intestinalis*, and zebrafish, which require that the theory described above be adapted and modified.

One variation includes the *distal-less* gene found in *Drosophila* which appears to have undergone a disassociation with the homeotic genes. In this case, the *Dll* gene is not found on the same chromosome as the homeotic genes. It has been suggested that since the homeotic genes themselves have become separated and do not appear in a cluster as seen in vertebrates, that this split occurred after the divergence of arthropods and vertebrates. The presence of a single *Dlx* gene in *C. intestinalis* that is unlinked to a *Hox* cluster, in addition to the bigene cluster, suggests another duplication event. This may have been another tandem duplication resulting in a single *Dlx* gene or an entire genomic duplication resulting in two linked *Dlx* genes followed by the loss of one of the two genes (Stock, Ellies et al. 1996). A similar situation is seen in the zebrafish. While this species contains three sets of *Dlx* gene pairs linked to unique *Hox* clusters, it also shows two additional *Dlx* genes, one of which is linked to a *Hox* cluster while the other is in a region of conserved synteny where it is believed another *Hox* cluster once existed (Amores, Force et al. 1998). It is believed that another duplication event, followed by degeneration of one of the duplicated genes, occurred

in the ray-finned fish after their divergence from the lineage of lobe-finned fish leading to tetrapods, resulting in the additional *Dlx* genes seen in zebrafish and other ray-finned fish (Stock, Ellies et al. 1996; Amores, Force et al. 1998; Force, Lynch et al. 1999; Sumiyama, Irvine et al. 2003).

1.7 Expression of the *Dlx* Genes

The expression domains of the *Dlx* genes in mice are primarily ectodermal derivatives, such as areas of surface ectoderm and parts of the nervous system. Secondary sites of *Dlx* expression include mesodermally derived cells, including skeletal tissue such as teeth, and the olfactory placodes and otic vesicles (Ellies, Stock et al. 1997; Zerucha and Ekker 2000; Panganiban and Rubenstein 2002). Many of these expression domains are conserved in other vertebrates, such as zebrafish, chickens and frogs. The expression domains are also common among the *Dlx* genes themselves. For example, the expression patterns of *Dlx3* and *Dlx4* are indistinguishable in the developing zebrafish, being expressed in the olfactory placodes, visceral arches, otic vesicles, and fins (Ellies, Stock et al. 1997). Similarly, *Dlx1*, *Dlx2*, *Dlx5* and *Dlx6* all show similar areas of expression in the developing forebrain (Ellies, Stock et al. 1997; Eisenstat, Liu et al. 1999).

1.7.1 *Dlx* in the Brain

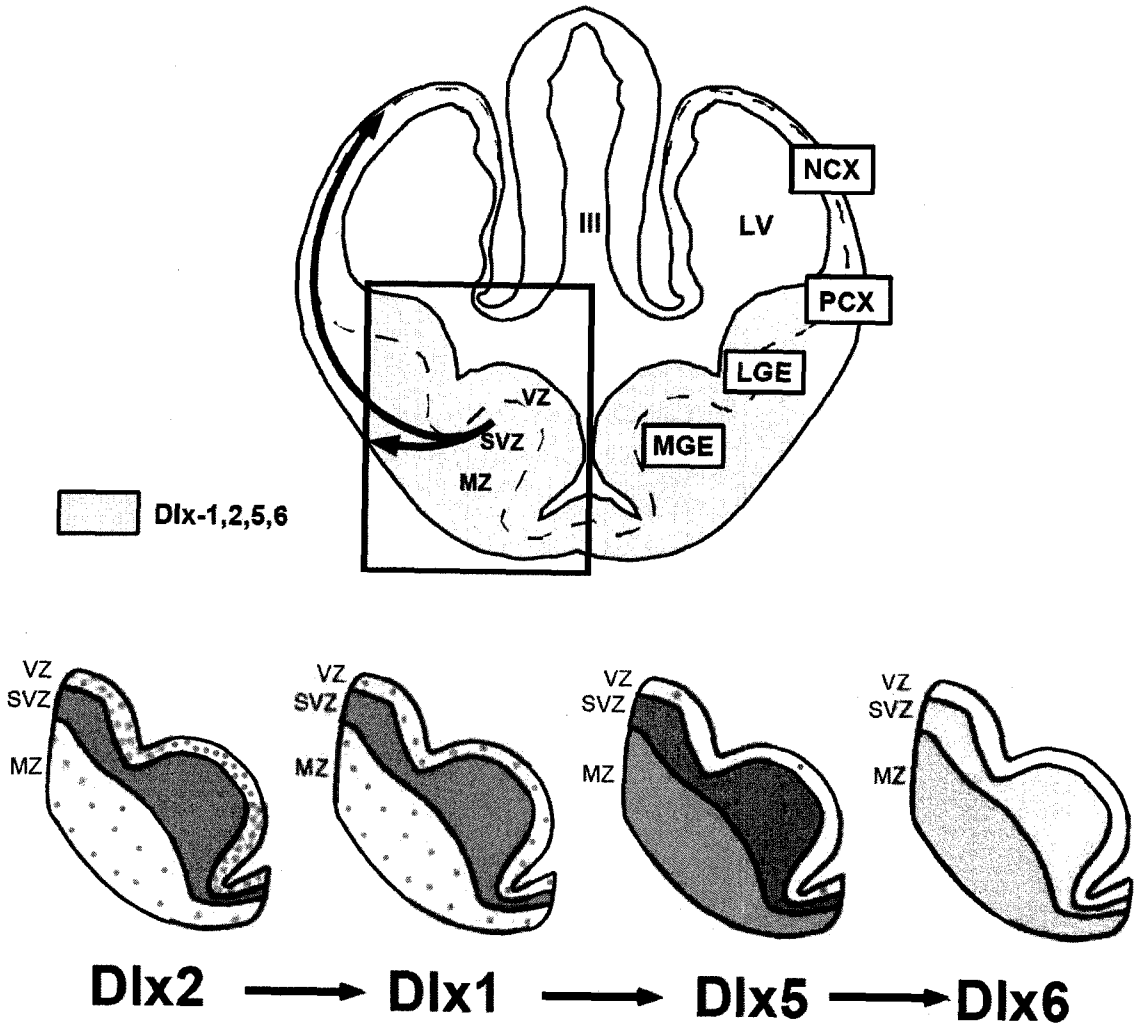
One of the major expression domains of the *Dlx* genes is within the central nervous system. Four of the six murine *Dlx* genes, *Dlx1*, *Dlx2*, *Dlx5*, and *Dlx6*, are expressed in areas of the forebrain (Bulfone, Kim et al. 1993; Liu, Ghattas et al. 1997; Eisenstat, Liu et al. 1999; Panganiban and Rubenstein 2002). This expression is divided into two regions, the telencephalon and the diencephalon, a pattern which is seen in other vertebrates (Panganiban and Rubenstein 2002). In the telencephalon, expression of the *Dlx* genes is restricted to

neural progenitors cells which will differentiate into GABAergic neurons and interneurons. Individual differentiating GABAergic neurons express each of the four forebrain *Dlx* genes, but the timing and location of the expression of these genes is unique for each of the genes (Anderson, Eisenstat et al. 1997; Stuhmer, Puelles et al. 2002). Cells first begin to express *Dlx2* during neurogenesis, which then results in the subsequent expression of *Dlx1*, *Dlx5* and then *Dlx6*. This temporal cascade of *Dlx* expression occurs as the cells of the developing forebrain mature, differentiate and migrate, with more mature cells expressing the later *Dlx* genes in areas of the forebrain closer to the mantle (Figure 1.2) (Ellies, Stock et al. 1997; Liu, Ghattas et al. 1997; Eisenstat, Liu et al. 1999; Panganiban and Rubenstein 2002). It is believed this regulatory cascade plays a fundamental role in determining the identity of GABAergic neurons and for directing their migratory pattern.

1.7.2 *Dlx* in the Branchial Arches

The cells of the branchial arches are populated by a subset of neural crest cells. These cells, and the structures they lead to, express all of the murine *Dlx* genes. These structures include the teeth, mouthparts, and skeletal and connective components of the developing face (Zerucha and Ekker 2000; Panganiban and Rubenstein 2002). As with the expression patterns in the brain, the branchial arches show a significant amount of overlap in the expression domains of the *Dlx* genes. The *Dlx* genes are also expressed in a spatial and temporal cascade, similar to that seen in the forebrain. It is believed that these genes help pattern the proximo-distal axis since the genes being expressed and their level of expression is dependent on their location along this axis. Within the branchial arches, *Dlx1* and *Dlx2* are expressed proximally, while *Dlx1*, *Dlx2*, *Dlx5* and *Dlx6* are expressed along the intermediate regions, and all six *Dlx* genes are expressed in the distal areas of the first and second

Figure 1.2. Areas of *Dlx* expression in the developing mouse forebrain. The migratory pathway of developing neurons is shown by the arrow. Specific domains of expression of *Dlx1*, *Dlx2*, *Dlx5*, and *Dlx6* are shown below. The number of cells expressing the *Dlx* gene in each region is demonstrated by the colour (uniform expression = dark colour; intermediate expression = light colour; scattered expression = dots). The spatial cascade of *Dlx* genes is demonstrated by different colours and the temporal cascade is illustrated below. NCX = cortex, PCX = palliocortex, LGE = lateral ganglionic eminence, MGE = medial ganglionic eminence, VZ = ventricular zone, SVZ = subventricular zone, MZ = mantle zone, LV = lateral ventricle, III = third ventricle of diencephalon (Adapted from Panganiban and Rubenstein 2002)


III
LV
NCX
PCX
LGE
MGE
VZ
SVZ
MZ
Dlx-1,2,5,6
VZ
SVZ
MZ
VZ
SVZ
MZ
VZ
SVZ
MZ
VZ
SVZ
MZ
Dlx2
Dlx1
Dlx5
Dlx6

branchial arches (Qiu, Bulfone et al. 1997; Zerucha and Ekker 2000; Beverdam, Merlo et al. 2002; Depew, Lufkin et al. 2002; Panganiban and Rubenstein 2002).

1.7.3 *Dlx* in Other Domains

The *Dlx* genes are also expressed in various elements of the surface ectoderm. In zebrafish and mouse, *Dlx* genes are expressed in the olfactory placodes and otic vesicles, (Ellies, Stock et al. 1997; Panganiban and Rubenstein 2002). All of the *Dlx* genes are expressed in the developing fins in zebrafish, homologous to the expression seen in the apical ectodermal ridge of developing limbs in mice and reflective of the role of *Dll* in the development of the *Drosophila*'s distal appendages (Dolle, Price et al. 1992; Ellies, Stock et al. 1997; Zerucha and Ekker 2000). Certain *Dlx* genes have also been shown to be expressed in other structures, such as the retina (*Dlx1* and *Dlx2*) and the mesodermally-derived hematopoeitic cells (*Dlx4*) (Panganiban and Rubenstein 2002).

1.7.4 Phenotypes of *Dlx* Mutants

The importance of *Dlx* gene expression for proper development has been demonstrated through experiments eliminating one or more of the *Dlx* genes. Mice homozygous for mutations in *Dlx1* or *Dlx2* lead to significant abnormalities in the branchial arches and their derivatives. Mutants are viable at birth but die within hours (*Dlx1/Dlx2*; *Dlx2*) or within the first month (*Dlx1*). Mice do not show a clear external phenotype for the mutation; however, abnormalities can be seen internally with alterations in the formation of the craniofacial bones in the proximal region of the first and second arches as well as the nerves associated with these structures. In *Dlx2* mutants, abnormal differentiation of the cells within the forebrain is evident. The abnormalities are generally restricted to the proximal structures, suggesting that other distally expressed *Dlx* genes compensate for the loss of *Dlx1*

or *Dlx2* in these regions (Qiu, Bulfone et al. 1995; Qiu, Bulfone et al. 1997). Mice homozygous mutant for the *Dlx5* gene show more severe phenotypic abnormalities. These include severe craniofacial defects and exencephaly (focused on the midbrain but can include the forebrain and hindbrain). Defects in the olfactory placode, otic capsule, and branchial arch derivatives are also seen in *Dlx5* mutants (Acampora, Merlo et al. 1999; Depew, Liu et al. 1999; Long, Garel et al. 2003). *Dlx3* null mutations also result in lethal phenotypes in mice. These embryos die around E10 due to failure of the placenta (Morasso, Grinberg et al. 1999). To date, *Dlx4* mutants have not been studied.

The overlapping pattern of *Dlx* gene expression, seen among many of the genes in many structures, suggests that there is a degree of redundancy among the *Dlx* genes. Mice mutant for *Dlx* genes have been used to demonstrate the level of this redundancy. Homozygous mutants for both *Dlx1* and *Dlx2* show more pronounced phenotypes than mice mutant for either *Dlx1* or *Dlx2*. This includes the complete loss of maxillary molars and loss of cell migration from the telencephalon to the neocortex, resulting in a massive reduction in GABAergic interneurons. Furthermore, the loss of both *Dlx1* and *Dlx2* results in the loss of nearly all *Dlx5* and *Dlx6* transcripts, fundamentally eliminating the expression of all four of the forebrain *Dlx* genes (Anderson, Eisenstat et al. 1997; Qiu, Bulfone et al. 1997; Zerucha and Ekker 2000).

Mice lacking expression of both *Dlx5* and *Dlx6* also show extreme phenotypes. These mice show a transformation of the lower jaw into an upper jaw (Figure 1.3). This is evidence of the proximalization of a distal structure (the lower jaw) after loss of the distally expressed genes (*Dlx5* and *Dlx6*). Although not conclusively shown, it has been suggested that the more distally expressed *Dlx5* and *Dlx6* genes compensate for the loss of the more proximally

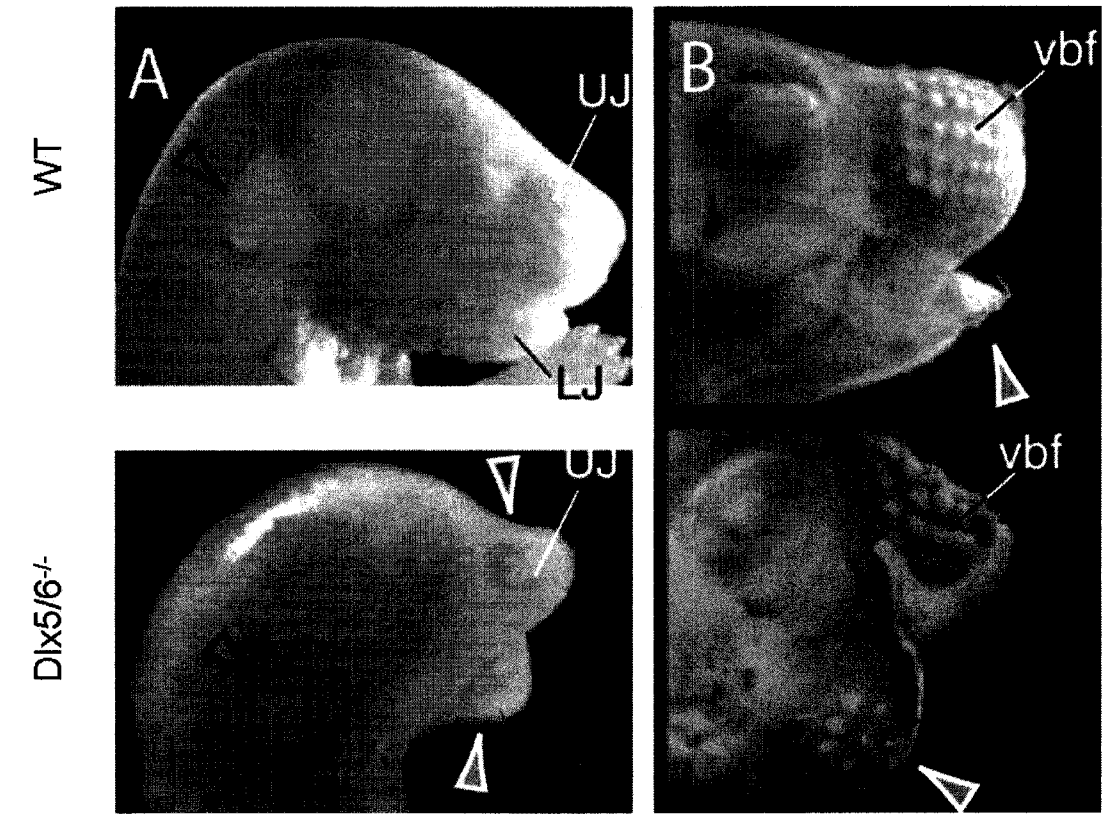
expressed *Dlx1* and *Dlx2* genes. Thus, abnormalities are only seen in the proximal structures, such as the bone abnormalities in *Dlx1/Dlx2* double mutants. In the *Dlx5/Dlx6* double-mutant, the reverse does not hold true in that the earlier and more proximally expressed *Dlx1* and *Dlx2* genes cannot compensate for the loss of *Dlx5* and *Dlx6*. Further, the most distally expressed *Dlx* genes, *Dlx3* and *Dlx4* are not able to compensate for the loss of *Dlx5* and *Dlx6*. This results in a more severe breakdown in the patterning of the proximo-distal axis and a more extreme phenotype in the *Dlx5/Dlx6* double mutants (Depew, Lufkin et al. 2002; Depew, Simpson et al. 2005).

A mutation removing the *dlx3* and *dlx4* genes in zebrafish has been studied; however, this mutation also results in the loss of expression of other genes (Solomon and Fritz 2002). These mutant fish fail to develop somites as well as otic and olfactory placodes, areas of endogenous *dlx* expression. This is also a lethal mutation and embryos die around 56 hours. Although these mutations are consistent with the effects of the loss of the *Dlx* genes, the phenotype cannot be definitively attributed to *Dlx* due to the simultaneous loss of other genes.

1.8 *Dlx* Regulation and DLX Targets

The *Dlx* genes encode homeobox transcription factors and are involved in a number of highly complex developmental pathways. Since many factors may be involved in these pathways, it has been difficult to specifically identify those that directly regulate *Dlx*. Retinoic acid, bone morphogenetic protein (BMP) and fibroblast growth factor (FGF) have been suggested as possible direct regulators of *Dlx*, but the nature of the interaction has not been established. Retinoic acid has been shown to rapidly reduce *Dlx* expression in zebrafish embryos, resulting in craniofacial abnormalities. *Dlx* has also been shown to be positively

Figure 1.3. Comparison of jaw formation in wild-type (WT) and *Dlx5/6*$^{-/-}$ mutant mouse neonates. A) Transformation of the lower jaw (LJ) seen in wild-type (top) into an upper jaw (UJ) in the absence of *Dlx5* and *Dlx6* (bottom, white and pink arrowhead). Note also the loss of external ear pinnae (white and red arrowhead) and nasal capsule elaboration (white and green arrowhead) in the *Dlx5/6*$^{-/-}$ mutant B) Removal of surface ectoderm reveals the presence of vibrissae (vbf) on upper jaw of both wild-type (top) and *Dlx5/6*$^{-/-}$ mutant mice but ectopic expression of vibrissae on lower jaw of mutant mouse (white and pink arrowhead), demonstrating the transformation of the lower jaw into an upper jaw. (Adapted from Depew, Simpson et al. 2005).


WT
Dlx5/6-/-
A
UJ
LJ
B
vbf
UJ
vbf

regulated by BMP and negatively regulated by FGF proteins but their mode of regulation is still unknown (Ellies, Langille et al. 1997; Zerucha and Ekker 2000). Interestingly, these same factors have been shown to have the opposite effect on *Dlx* intergenic enhancers (discussed below) with FGF showing a positive response and BMP-4 having a negative action (Park, Sperber et al. 2004).

The formation of protein homodimers and heterodimers with *Dlx* have also been shown to function in the regulation of the *Dlx* genes. DLX proteins are able to form dimers with MSX proteins (Zhang, Hu et al. 1997; Bendall and Abate-Shen 2000). *Msx* is expressed in similar domains as *Dlx*, such as the neural tube, limb buds and teeth, and the genes share 58% similarity in their homeodomain. However, MSX proteins act as transcriptional repressors, unlike DLX proteins which generally function as transcriptional activators. In the dimeric form, the binding activity of both the MSX and DLX proteins is reduced. As such, MSX regulates the activational activity of DLX by binding to the protein. Similarly, DLX targets MSX and reduces its activity as a repressor (Zhang, Hu et al. 1997; Bendall and Abate-Shen 2000).

Masuda et al. have identified a novel protein, DLXIN-1, or MAGE-D1, that binds DLX5. The mRNA for this protein is found in a wide variety of adult and embryonic tissue, including bone. DLXIN-1 has been shown to activate the transcriptional activity of DLX5 (Masuda, Sasaki et al. 2001). Additionally, Kuwijima et al. have shown that necdin promotes the differentiation of GABAergic neurons in the forebrain by binding DLX via DLXIN-1 (Kuwajima, Nishimura et al. 2006).

As with the identification of factors regulating *Dlx*, the number of identified direct targets of *Dlx* is minimal. One clear direct *Dlx* target is the Neuropilin-2 (NRP-2) receptor

(Le, Du et al. 2007). The NRP-2 promoter contains seven nucleotides of the core DLX consensus DNA binding sequences (5'-ATAATTA-3'). Both DLX1 and DLX2 are able to inhibit NRP-2 expression. This is supported by the fact that *Dlx1/Dlx2* null mice show increased NRP-2 expression in the forebrain. Since NRP-2 functions as an inhibitor of neuronal migration, by inhibiting NRP-2 expression, DLX1 and DLX2 facilitate neuronal migration, specifically in the forebrain.

Although the majority of studies have shown the *Dlx* genes to function as transcriptional activators, evidence of transcriptional repression via *Dlx* genes has also been found. DLX5 has been implicated in the repression of *osteocalcin* gene transcription. The *osteocalcin* gene has a role in bone formation and growth during development. It has been suggested that hyperactivity of this gene prevents proper bone formation. This may account for the skeletal malformations seen in the development of *Dlx* mutant mice (Ryoo, Hoffmann et al. 1997).

Cobos et al. used gain-of-function and loss-of-function experiments to show the role of *Dlx* genes in regulating the *Aristaless* (*Arx*) gene (Cobos, Broccoli et al. 2005). The *Arx* gene is expressed in areas of the forebrain in which *Dlx* is also expressed and has a role in the migration of GABAergic neurons, as is also seen in the *Dlx* genes. The study provided evidence that ectopic expression of *Dlx* in neuronal cells induces *Arx* expression. Furthermore, mutant embryos lacking DLX1 and DLX2 show a reduction in ARX. This suggests a role of DLX in the regulation of the *Arx* gene. Since the *Drosophila Aristaless* gene is downstream of the *Dll* gene, this could represent a conserved regulatory function of the vertebrate *Dlx* genes. *Dlx* has been implicated in many other pathways but the direct targets of the DLX proteins have not been identified.

Finally, another type of DLX target that has been suggested are the highly conserved, non-coding sequences located within the intergenic regions of the *Dlx* genes. Given the organization of the *Dlx* genes into closely linked bigene clusters and the overlapping expression patterns seen among the genes, it has been suggested that *cis*-acting regulatory elements found in the intergenic regions may be playing a role in controlling the expression of these genes (Ellies, Stock et al. 1997). These sequences may be involved in auto-regulatory mechanisms for DLX, in the regulation of other *Dlx* genes, or the regulation of other genes altogether. Confirming this theory, a number of *cis*-acting regulatory elements have been found in the intergenic regions of all three vertebrate bigene clusters.

1.9 *Cis*-Acting Regulatory Elements

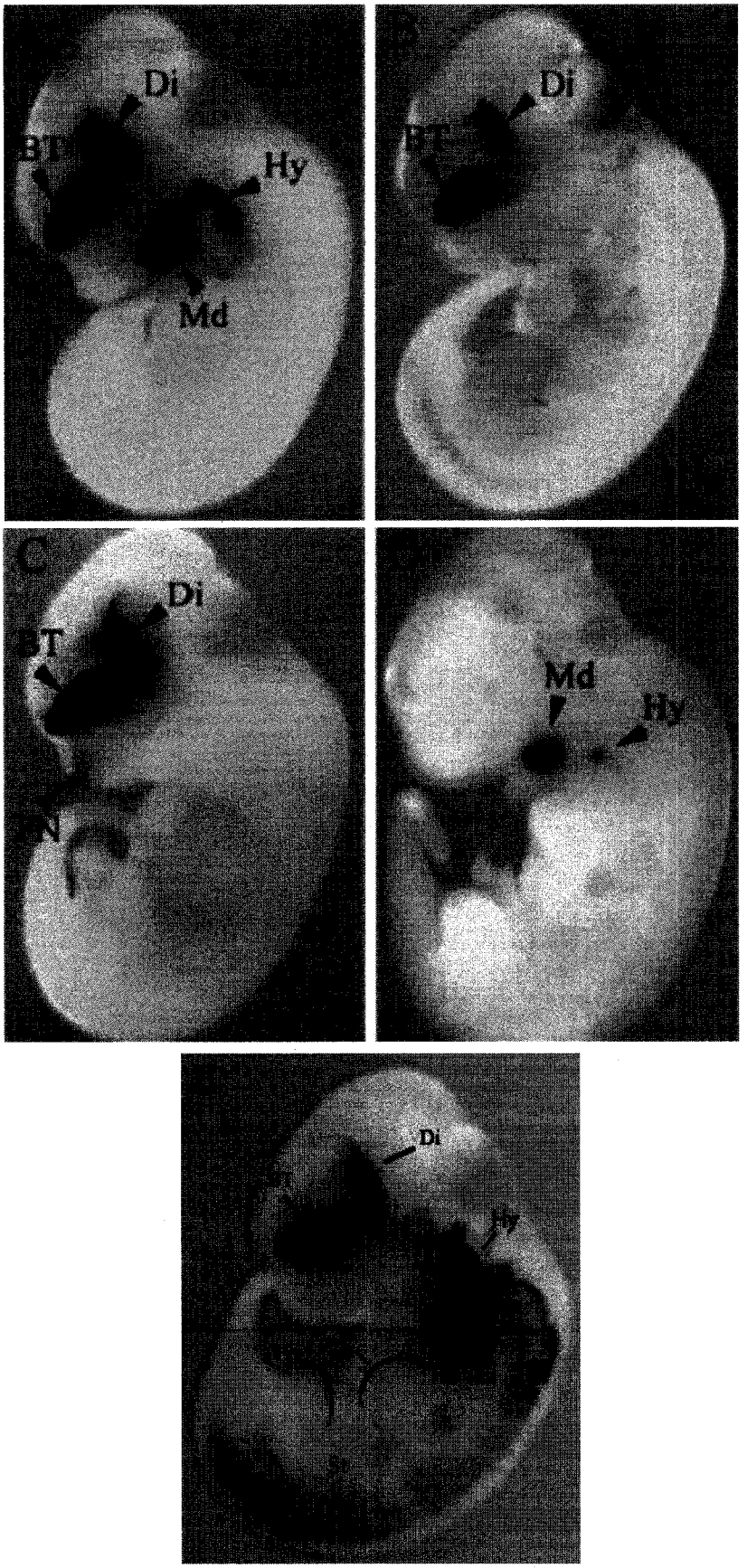
Several *cis*-acting regulatory elements have been identified within the loci of each of the three vertebrate *Dlx* bigene clusters (Zerucha, Stuhmer et al. 2000; Ghanem, Jarinova et al. 2003; Sumiyama and Ruddle 2003; Ghanem, Yu et al. 2007). The I56i and I56ii elements have been identified in the *Dlx5/Dlx6* locus. The I56ii sequence is approximately 300 base pairs in length and is located in the *Dlx5/Dlx6* intergenic region, closer to the *Dlx5* gene when compared to I56i. It is 79-99% conserved between human, mouse, zebrafish and the two species of pufferfish. Activity is seen in both the telencephalon and the diencephalon, as well as the apical ectodermal ridge (AER) (Figure 1.4 B). The I56i enhancer, which is the focus of this research, is discussed in detail below.

Within the *Dlx1/Dlx2* locus, the I12a, I12b, and URE2 elements have been studied in detail. The I12a element is highly conserved (83-99% among humans, mice, zebrafish and two species of pufferfish) and is active in the mesenchyme of the first and second branchial arches (Figure 1.4 D) (Ghanem, Jarinova et al. 2003). The I12b enhancer is 75-97% among

the species mentioned above. Its activity is seen in the telencephalon and diencephalon, as well as the AER (Figure 1.4 C). It also shows activity in the frontonasal prominence (Ghanem, Jarinova et al. 2003). There are a number of similarities found between I12b and the I56i element, discussed in detail below. The URE2 element is located approximately 12kb upstream of the *Dlx1* start site. The activity of this element is seen in the telecephalon and diencephalon, as well as the branchial arches, somites and AER (Figure 1.4 E) (Ghanem, Yu et al. 2007).

The elements found in the *Dlx3/Dlx4* intergenic region show less similarity (85-90%) than the enhancers discussed above, when comparing mouse and human sequences. Individual enhancer studies have not been reported, but the locus containing a number of enhancers from this region shows activity in the AER and the branchial arches. No activity in the forebrain has been reported (Sumiyama and Ruddle 2003).

Figure 1.4. Activity of the five identified cis-acting regulatory elements in the *Dlx1/Dlx2* and *Dlx5/Dlx6* loci as seen in E11.5 mouse embryos carrying *lacZ* reporter constructs A) I56i B) I56ii C) I12B D) I12A E) URE2. Activity can be seen in two domains of the forebrain, the basal telencephalon (BT) and the diencephalon (Di), and the first two branchial arches, the mandibular (Md) arch and the hyoid (Hy) arch. Other areas of activity include the frontonasal prominence (FN), AER of the limbs (L) and the somites (St). (Adapted from Ghanem, Jarinova et al. 2003 (A-D) and Ghanem, Yu et al. 2007 (E)).


Di
Hy
Md
Di
Di
Md
Hy
Di
Hy

1.10 The I56i Enhancer

1.10.1 Identification of the I56i Element

The I56i enhancer is one of two *cis*-acting regulatory elements identified in the intergenic region of the *Dlx5/Dlx6* locus. The I56i sequence is 400 base pairs (bp) in length and is located closer to the *Dlx6* gene in both mouse and zebrafish (Zerucha, Stuhmer et al. 2000). The I56i sequence is 81-99% conserved when comparing human, mouse, zebrafish and two species of pufferfish, *Spheroides nephelus* and *Takifugu rubripes* (Ghanem, Jarinova et al. 2003). In fact, the I56i element is ultraconserved (100% identity over at least 200 base pairs) between mouse, rat and human (Bejerano, Pheasant et al. 2004). This high level of sequence conservation suggests that this element is fundamental in the developmental pathways of many diverse organisms.

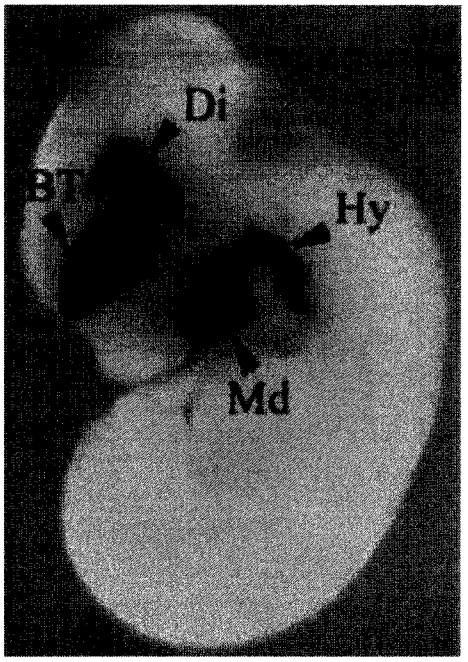
1.10.2 Activity of the I56i Element

The activity of the enhancer has been analyzed using a *lacZ-β-globin* minimal promoter reporter construct (Zerucha, Stuhmer et al. 2000; Ghanem, Jarinova et al. 2003). It was observed that the activity of the I56i enhancer in both mouse and zebrafish recapitulates many of the areas of endogenous *Dlx* expression (Figure 1.5). I56i shows activity in two regions of the forebrain, in the telencephalon and the diencephalon. In the mouse, activity of the I56i enhancer is also seen in the branchial arches, olfactory placodes and the AER, where endogenous *Dlx* expression is also observed. It is active beginning at embryonic day 10 and continues past P35 (Ghanem, Yu et al. 2007).

1.10.3 I56i Activity in *Dlx* Mutants and Possible Role in Cross-Regulation

As previously discussed, the loss of *Dlx1* and *Dlx2* results in a significant reduction in the expression of *Dlx5* and *Dlx6*, resulting in defective neurogenesis in the developing

Figure 1.5. Activity of the I56i enhancer in an E11.5 mouse embryo carrying a *lacZ* reporter construct. Activity is seen in two elements of the forebrain, the basal telencephalon (BT) and the diencephalon (Di). Activity is also seen in the first two branchial aches, the mandibular (Md) and hyoid (Hy) arches. (Adapted from Ghanem, Jarinova et al. 2003).


Di
BT
Hy
Md

forebrain. Zerucha et al. demonstrated that the activity of the I56i enhancer is significantly reduced in mice homozygous for the deletion of both *Dlx1* and *Dlx2* (Zerucha, Stuhmer et al. 2000; Ghanem, Jarinova et al. 2003). They suggest that the I56i enhancer has a role in regulating the expression of *Dlx5* and *Dlx6* via transcriptional activation of the I56i enhancer through *Dlx1* and/or *Dlx2*.

The I56i enhancer also contains two putative DLX binding sites, supporting their theory. Mutagenesis of either or both of the DLX binding sites in the I56i enhancer results in a significant reduction in the activity of the I56i enhancer in the forebrain and the ability of DLX2 to activate transcription via the enhancer. These results support the idea that activity of the I56i enhancer in the forebrain is dependent on activation via DLX1 or DLX2. Furthermore, DNA binding assays demonstrate that DLX2 binds a fragment of I56i containing the two putative *Dlx* binding sites. This is reduced following mutagenesis of either of the two sites and lost following mutagenesis of both sites (Zerucha, Stuhmer et al. 2000). Research by Zhou et al. confirms the binding of *Dlx2* to these two putative DLX binding sites within the I56i enhancer (Zhou, Le et al. 2004). Furthermore, they demonstrate that DLX1 is also able to bind the same two regions of the enhancer. The ability of DLX2 to activate transcription was also confirmed via a luciferase reporter gene. Mutagenesis of the two putative DLX binding domains within the I56i enhancer again resulted in a reduction in reporter activity, confirming that DLX2 interacts with the I56i enhancer to activate transcription.

1.10.4 Identification of *Trans*-Acting Factors Regulating I56i Activity in the Branchial Arches

As mentioned above, many of the factors that have been shown to influence expression of the *Dlx* genes have also been shown to affect the activity of the *Dlx* intergenic enhancers, including I56i. In mice, Park et al. demonstrated that removal of the mandibular epithelium reduced *lacZ* reporter activity in construct carrying the I56i enhancer (Park, Sperber et al. 2004). They were further able to identify a number of *trans*-acting factors present in the epithelium that affected the activity of these reporters. Both FGF8 and FGF9 induced I56i reporter activity in areas of endoegenous I56i activity. Exposure to BMP4 counteracted this increase in reporter activity and further reduced I56i reporter activity. Exposure to bosentan, an inhibitor to endothelin-1 (ET-1), again resulted in a reduction in I56i reporter activity. However, this reduction was restricted to distal areas of the branchial arches and is similar to the reduction in *Dlx6* expression seen in mice deficient of the type A endothelin receptor. Exposure to FGF8, FGF9, and BMP4 resulted in similar changes in reporter activity in a construct carrying the I12a enhancer element. As mentioned above, these epithelial factors, such as FPF8 and BMP4, have been shown to regulate endogenous *Dlx* expression. This highlights the regulatory role of two of the *Dlx* intergenic enhancers in mediating *Dlx* expression via a number of *trans*-acting factors present in the branchial arches.

1.10.5 I56i as a Non-Coding RNA

Although the intergenic region between *Dlx5* and *Dlx6* is generally considered to be an un-transcribed sequence, the I56i enhancer has been shown to code for a portion of the second exon in *Evf-2*, a non-coding RNA (Feng, Bi et al. 2006). This RNA is detected in regions of the telencephalon, the lateral ganglionic eminence (LGE) and the medial ganglionic eminence (MGE), congruent with endogenous activity of the I56i enhancer. *Evf-2* is expressed in immature neurons migrating out of the ventricular zone. This study showed that *Evf-2*, as a single-stranded sense RNA, increased reporter activity in the presence of DLX. This occurred via both enhancers in the *Dlx5/Dlx6* intergenic region, though greater activation is seen through the I56i enhancer. Cooperation with DLX2 resulted in the greatest increase in reporter activity, while other forms of DLX *were* also able to increase activity. Overall, this study suggests that the I56i enhancer is not only significant as a *cis*-acting regulatory element but also plays a role in regulating *Dlx5* and *Dlx6* as a transcription-regulating, ultraconserved ncRNAs.

1.10.6 I12b – A *Dlx* Intergenic Enhancer Similar to I56i

A *cis*-acting regulatory element located in the intergenic region of the *Dlx1/Dlx2* locus shows some similar characteristics to the I56i enhancer. The I12b enhancer is 75-97% conserved among human, mouse, zebrafish and two species of pufferfish. Like the I56i enhancer, its activity is also seen in the telencephalon and diencephalon, as well as the AER. It also shows activity in the frontonasal prominence (Ghanem, Jarinova et al. 2003) (Figure 1.4 C). Despite the fact that I12b and I56i are both highly conserved, there is little sequence similarity between the two elements. One exception is the presence of two putatitive DLX

binding sites separated by similar distances. A putative homeodomain binding site (TAAT) is also found between the two DLX binding sites in both enhancers.

An analysis of the role of the I12b enhancer has previously been reported (Poitras, Ghanem et al. 2007). In this study, six putative binding sites were identified within the I12b enhancer element by *DNase*I footprinting. Two to nine base pair mutations were introduced into these sites. The effect of the base pair mutations on activity of the I12b enhancer was evaluated in mice using a *lacZ* reporter construct and evaluated *in vitro* by electrophoretic mobility shift assays.

It was shown that DLX2, and a number of other DLX proteins, were able to initiate transcription through the I12b enhancer. This occurs via the putatitve DLX binding sites, since mutagenesis of these sites eliminated transcriptional activity. It was also demonstrated that this mutagenesis resulted in reduced activity of the I12b enhancer in both the telencephalon and diencephalon (Figure 1.6). Other proteins, such as MASH1 and MEIS1, were also shown to bind to the I12b enhancer. Mutagenesis of the putative binding sites for both proteins showed decreases in enhancer activity in the transgenic embryos produced (Figure 1.6). Mutagenesis of the MASH1 putative binding site also prevented the binding of MASH1 proteins to the DNA of the I12b enhancer sequence (Figure 1.7). The results from this study demonstrate the factors affecting the activity of a *cis*-acting regulatory element and demonstrate a link between this activity and regulation of the *Dlx* genes. This provides support for the methods chosen for the research presented in this thesis and demonstrates the potential results that may be obtained.

Figure 1.6. Mutagenesis of putative transcription factor binding sites affects I12b activity in the forebrain. A) Activity of a *lacZ* reporter construct in E11.5 transgenic embryos carrying the wild-type I12b enhancer sequence (A) and I12b enhancer sequences with mutations introduced into each of six putative transcription factor binding sites (B-E). The mutated putative binding site is shown (HDTF = homeodomain transcription factor). Di = diencephalon; VT = ventral telencephalon; FN = frontonasal prominence. (Adapted from Poitras, Ghanem et al. 2007)

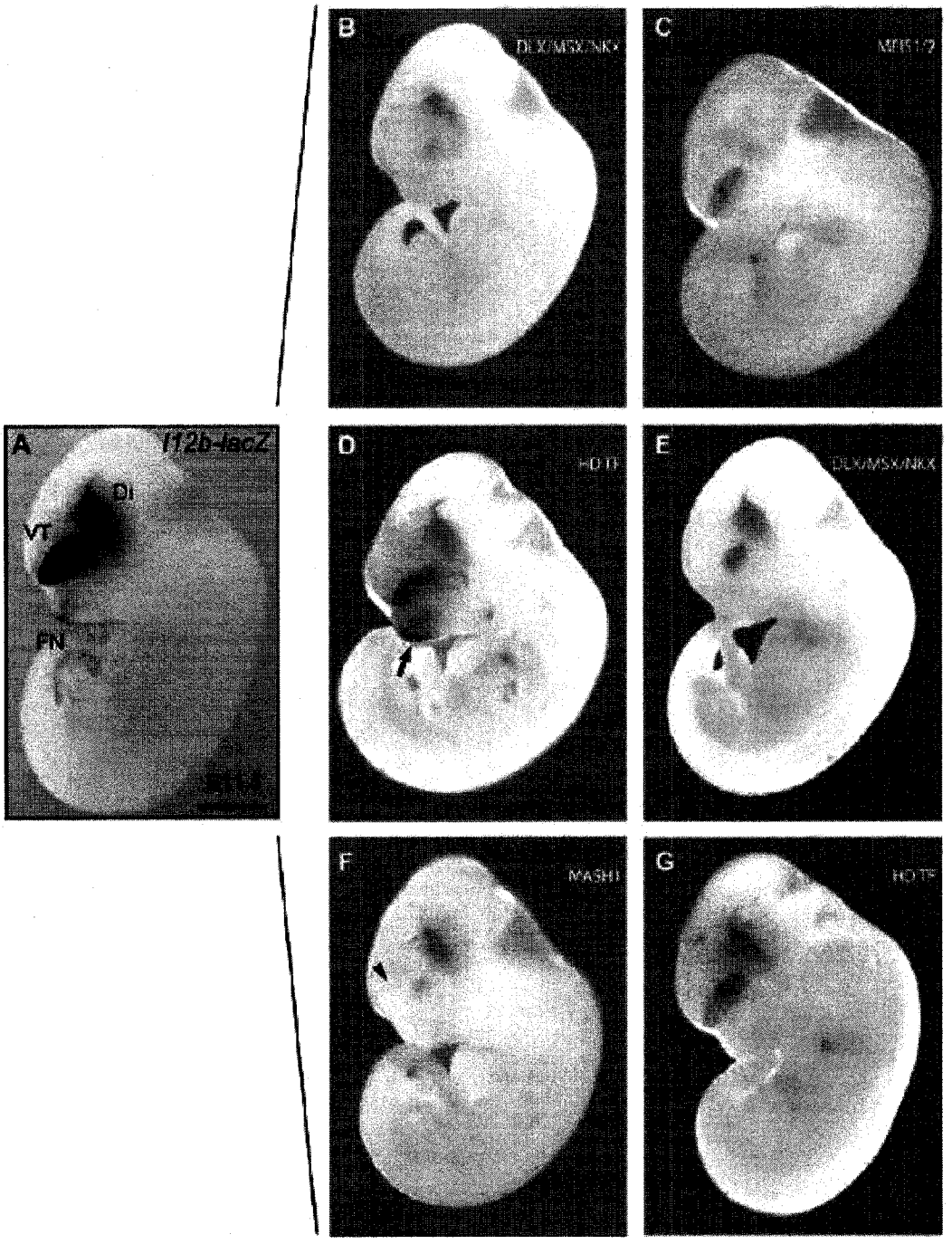
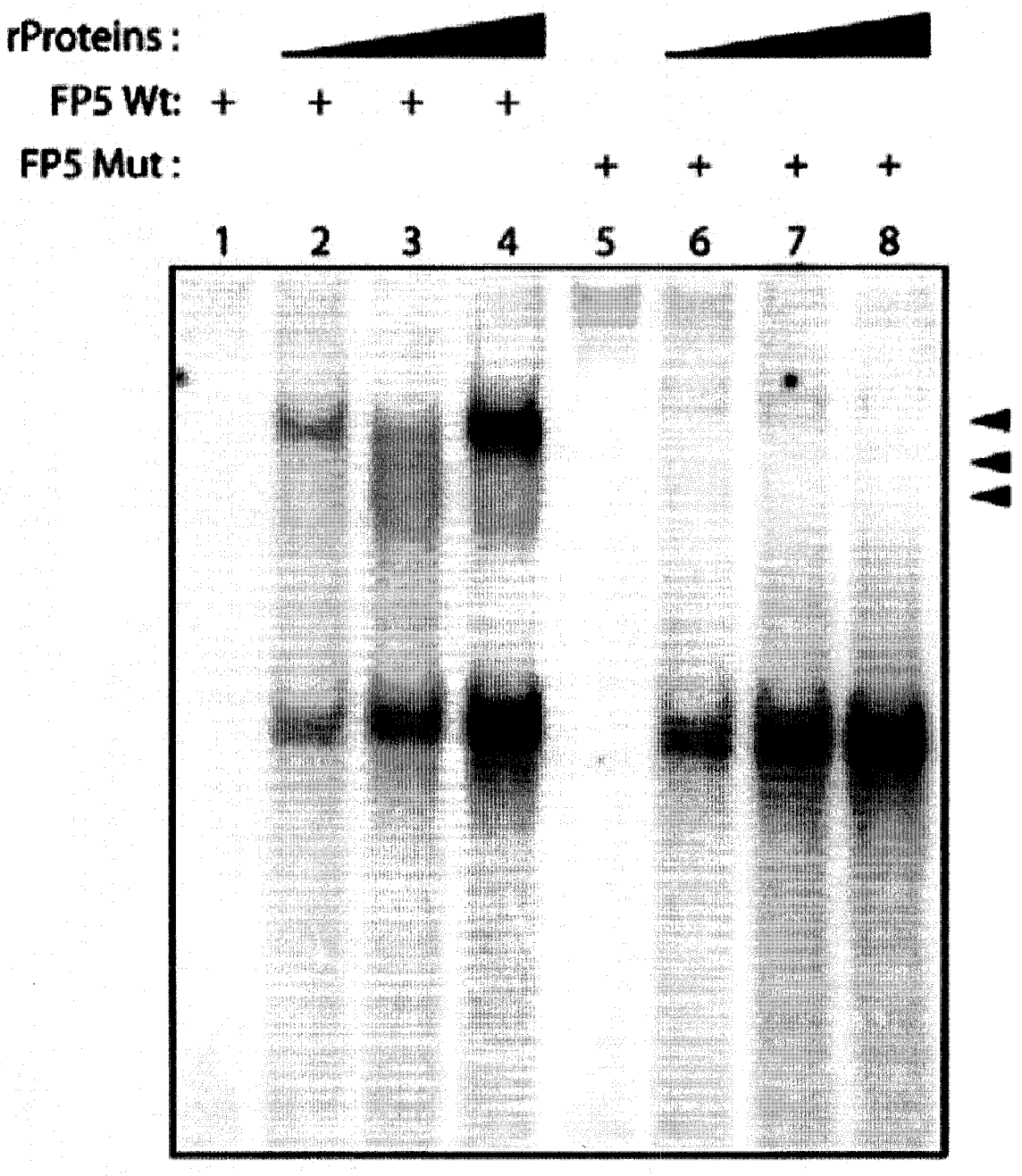

A
Tf2b-lacZ
Di
VT
PN
B
C
D
E
F
G

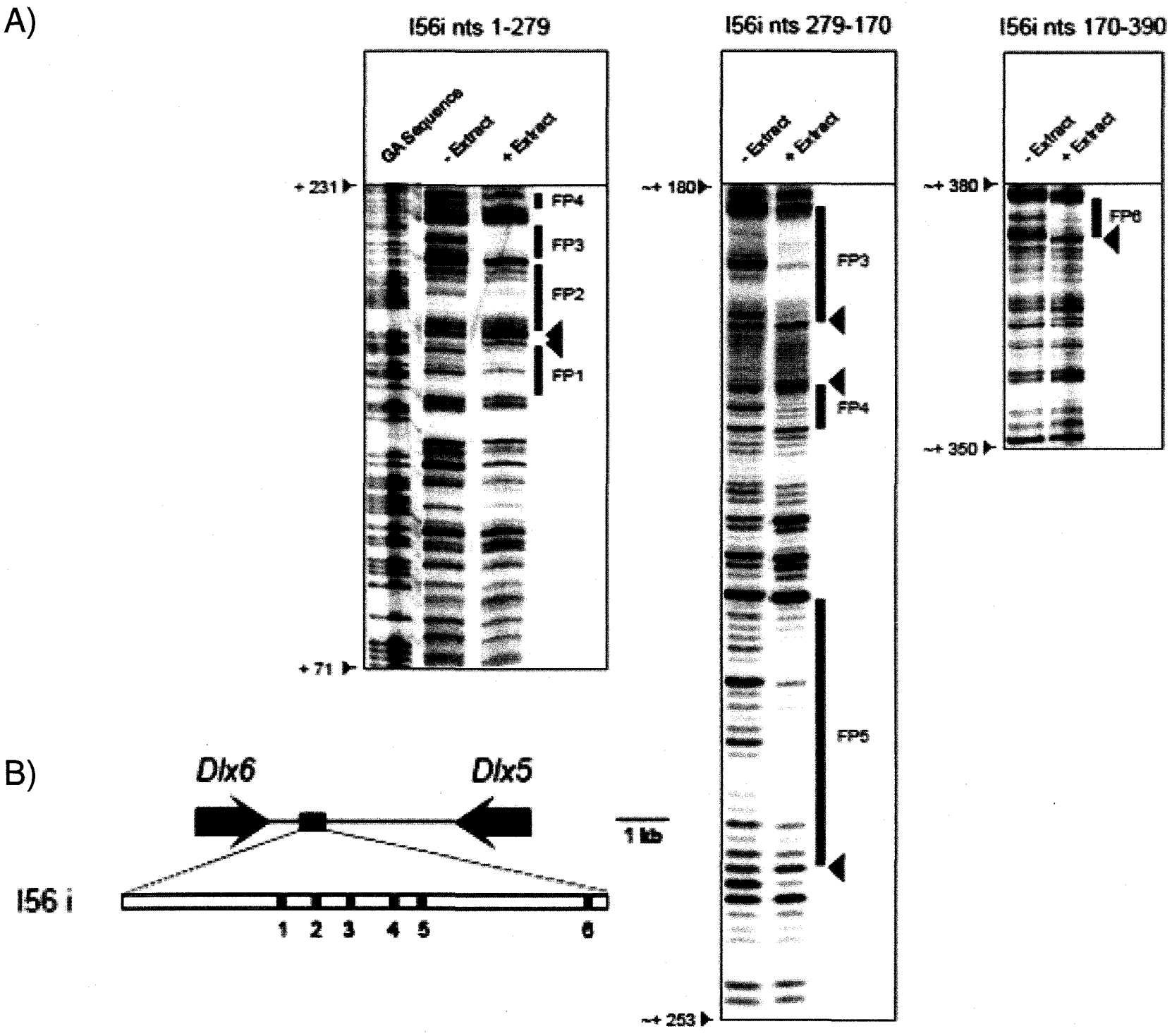
Figure 1.7. Mutation of a putative MASH1 binding sites prevent protein binding. EMSA demonstrates that the MASH1 protein, synthesized by *in vitro* transcription and translation (along with its cofactor, E47), binds a sequence of the I12b enhancer (lanes 2-4, arrows). Mutation of the putative binding site prevents protein binding (lanes 6-8). (Adapted from Poitras, Ghanem et al. 2007).


rProteins :
FP5 Wt: + + + +
FP5 Mut : + + + +
1 2 3 4 5 6 7 8

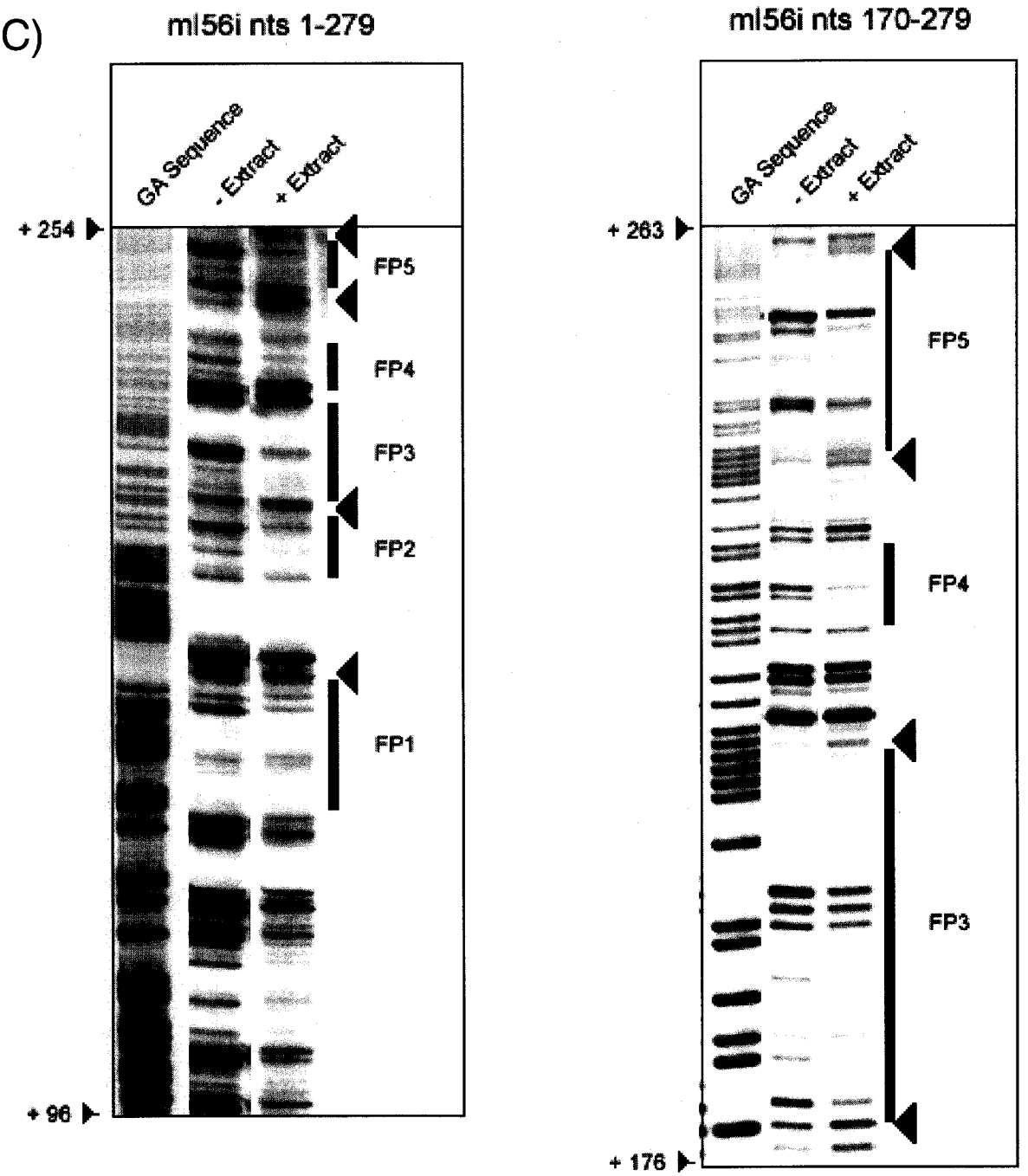
1.11 Initial Identification of Putative Transcription Factor Binding Sites within the I56i Enhancer

In a preliminary effort to identify the factors binding to the I56i enhancer, *DNase*I footprinting was previously completed on the I56i enhancer using nuclear extracts from the forebrains of E13.5 mouse embryos. Three overlapping fragments of 150-300 base pairs were used to cover the entire I56i enhancer. The I56i enhancer sequence for both zebrafish and mouse was used. From these experiments, six putative protein binding sites were identified, FP1-FP6 (Figure 1.8) (Luc Poitras, unpublished results).

Figure 1.8. *DNaseI* footprinting identifies six putative binding sites within the I56i enhancer. Proteins from a nuclear extract of E13.5 mouse forebrain bind three overlapping fragments of the zebrafish (a) and mouse (c) I56i enhancer at six sites, FP1-FP6. Thick lines represent protected regions and arrowheads indicate hypersensitive sites. Relative positions within the enhancer are given to the left of each figure. A schematic representation of the location of each of the putative binding sites within the I56i enhancer is shown in (b). (Luc Poitras, unpublished results)


A)
I56i nts 1-279
I56i nts 279-170
I56i nts 170-390
GA Sequence
- Extract
+ Extract
+ 231
+ 71
FP4
FP3
FP2
FP1
~+ 180
~+ 253
FP5
~+ 380
~+ 350
FP6
B)
Dlx6
Dlx5
1 kb
I56 i
1 2 3 4 5 6

C)


ml56i nts 1-279
GA Sequence
- Extract
+ Extract
+ 254
+ 96
FP5
FP4
FP3
FP2
FP1
ml56i nts 170-279
GA Sequence
- Extract
+ Extract
+ 263
+ 176
FP5
FP4
FP3

1.12 Objectives of this Project

As described above, previous studies have demonstrated the activity of the I56i enhancer in the forebrain and the branchial arches. The ability of the DLX proteins to bind this fragment of DNA has also been shown, as well as the effect of mutagenesis of one or both of the putative DLX binding sites found within the enhancer sequence. It is hypothesized that alteration of this element, such as seen with the mutagenesis of the DLX binding sites, will affect the ability of key proteins to bind to the enhancer. This will subsequently result in a change in the activity of the enhancer. The goal of this research is to further analyze the role of the I56i enhancer and its interactions with any proteins that may be involved in the regulation of the *Dlx* genes.

In order to achieve this goal, a number of objectives were devised.

- Identify regions within the enhancer that may potentially act as protein binding sites
- Evaluate the effect of mutagenesis of the identified putative binding sites on I56i enhancer activity
- Assess the ability of forebrain/brain proteins to bind the I56i enhancer
- Analyze the effect of mutagenesis of putative binding sites on the ability of forebrain/brain proteins to bind the I56i enhancer
- Identify the forebrain/brain proteins that may be binding the I56i enhancer

2. MATERIALS AND METHODS

2.1 Preparation of Nuclear Extracts

Forebrain tissue was dissected from E13.5 mouse embryos and brain tissue was dissected from mouse embryos harvested at E11.5. Nuclear extracts were obtained from these tissues, as described in Sambrook and Russell (Sambrook and Russell 2001). The extracts were then dialyzed in resuspension buffer in Spectra/Por CE Float-A-Lyzers (Spectrum Laboratories Inc.) overnight at 4°C. Aliquots of 20µL and 50µL were frozen in liquid nitrogen and stored at -80°C.

2.2 *DNase*I Footprinting Analysis

2.2.1 Preparation of DNA Fragments

One of three overlapping fragments of the I56i enhancer was amplified by PCR for use in *DNase*I footprinting. The primers used were 187.For and 436.Rev (Table 2.1). The conditions for the PCR were 1X PCR buffer (100mM Tris-HCl pH8.3, 500mM KCl, 15mM $MgCl_2$), 0.2mM dNTPs, 10-100ng template DNA, 0.2pM 187.For primer, 0.2pM 436.Rev primer and 1unit of Taq polymerase. The thermal cycling parameters were 1 cycle at 94°C for 5 minutes, 30 cycles of: 30 seconds at 94°C, 30 seconds at 50°C and 30 seconds at 72°C; followed by a final extension of 72°C for 10 minutes, and cooling at 4°C.

The PCR product was ligated into the pDrive vector (Qiagen) following manufacturer's protocols. The ligation product was transformed into XL-10 *Escherichia coli* chemically competent cells and grown in 1mL of luria broth (LB) for 1 hour at 37°C. These cells were plated on Ampicillin/Agar/LB plates containing IPTG (0.1M) and X-gal (0.09M) and grown overnight at 37°C. Positive clones were identified by blue/white

Table 2.1. Primers Used in All Experiments. The name of the primer, the primer sequences, and the experimental use are shown.

Primer Name	Primer Sequence	Location within I56i Enhancer	Experimental Use
187.For	CGG AAT TCT TTT TTC CAT TCT CTT AAA T	157-175	DNaseI Footprinting
436.Rev	CGG TAC CGC ATT ATA ATT TTG GTT AAT	380-400	DNaseI Footprinting
1.For	GAA AAA TGT TTT CTT TTC TTT T	1-22	Mutagenesis
84.For	CCT CCA GCT GCA GTG C	84-99	Mutagenesis
400.Rev	GCA TTA TAA TTT TGG TTA ATT T	378-400	Mutagenesis
Mut1.Rev	ACC TTC TTG TCT AGA AAA ACC TTA A	131-155	Mutagenesis
Mut2.Rev	TGG CTG CAT TTT CTA GAA TGG AAA AAA	156-182	Mutagenesis
Mut3.Rev	AAA TTA CTC TCT AGA TGG CTG CAT	174-197	Mutagenesis
Mut4.Rev	AAA ACG CTG TCT AGA GCG GGC TAC	203-226	Mutagenesis
Mut5.Rev	TTA CAG GTC TAG ATC TTT GAC	231-251	Mutagenesis
Mut6.Rev	GCT TTT TAG TCT AGA AAT TGG ACG ACA TCT AGA GGT AAT TAT	247-279	Mutagenesis
Mut7.Rev	CCT GTT CCT CTA GAT AAA TAA AGA	282-305	Mutagenesis
Mut8.Rev	CTC GAG GCA TCC CCA TTT TGG TCT AGA TTC AAT TTT A	370-400	Mutagenesis
24.For	GAA AAA TGT TTT CTT TT	1-17	EMSA
187.For	TTT TTT CCA TTC TCT TA	156-172	EMSA
316.Rev	GGT ACC GGC TTT TTA GTA TTA AA	264-280	EMSA
436.Rev	GGT ACC GCA TTA TAA TTT TGG TT	383-400	EMSA
1.19WT.For	TTA AGG TTT TTA TCC ACA AGA AGG TT	131-156	EMSA
1.19WT.Rev	AAC CTT CTT GTG GAT AAA AAC CTT AA	131-156	EMSA
1.19Mut.For	TTA AGG TTT TTC TAG ACA AGA AGG TT	131-156	EMSA

1.19Mut.Rev	AAC CTT CTT GTC TAG AAA AAC CTT AA	131-156	EMSA
2.7WT.For	TTT TTC CAT TCT CTT AAA TGC AGC CA	157-182	EMSA
2.7WT.Rev	TGG CTG CAT TTA AGA GAA TGG AAA AA	157-182	EMSA
2.7Mut.For	TTT TTC CAT TCT AGA AAA TGC CAG CCA	157-182	EMSA
2.7Mut.Rev	TGG CTG CAT TTT CTA GAA TGG AAA AA	157-182	EMSA
3.9WT.For	AAT GCA GCC ATA ATT AGA GTA ATT TT	173-198	EMSA
3.9WT.Rev	AAA ATT ACT CTA ATT ATG GCT GCA TT	173-198	EMSA
3.9Mut.For	AAT GCA GCC ATC TAG AGA GTA ATT TT	173-198	EMSA
3.9Mut.Rev	AAA ATT ACT CTC TAG ATG GCT GCA TT	173-198	EMSA
4.4WT.For	TGT AGC CCG CTG ATT ACA GCG TTT TT	202-227	EMSA
4.4WT.Rev	AAA AAC GCT GTA ATC AGC GGG CTA CA	202-227	EMSA
4.4Mut.For	TGT AGC CCG CTC TAG ACA GCG TTT TT	202-227	EMSA
4.4Mut.Rev	AAA AAC GCT GTC TAG AGC GGG CTA CA	202-227	EMSA
Bg.For	AGG GCA GAG CCA TCT ATT GC	n/a	Transgenic Screening
lacZ.Rev	CGC TCA TCC GCC ACA TAT CC	n/a	Transgenic Screening
FH.X1	GAT CAT GAC CGC CGT AGG	n/a	Transgenic Screening
FH.X2	CAT GAA CTT GTC CCA GGC TT	n/a	Transgenic Screening

Notes: Restriction sites integrated into the primer are underlined.

Base pair changes in the mutant primers are shown in red.

Base pair position 1 is defined as the base pair closest to the *Dlx6* gene.

selection, as clones containing an insert would disrupt the *lacZ* gene and thus disrupt the production of blue staining product. White colonies were selected and grown overnight at 37°C with shaking in 2mL of LB with 10μg/mL ampicillin overnight.

Plasmid DNA was extracted from the cultures using the Wizard Plus© Plasmid Purification Kit (Promega). Presence of the PCR product was verified through digestion with *Eco*RI. The positive clone was re-grown overnight in 100mL LB with 10μg/mL ampicillin. Plasmid DNA was extracted using the Qiagen Plasmid Midi Kit and the fragments were recovered with digestions of *Hind*III and *Pst*I followed by gel purification (QIAquick Gel Extraction Kit, Qiagen) following manufacturer's guidelines. Digestion with these enzymes ensured only one 5' overhang (*Hind*III) to allow for the targeted labelling of one end of the DNA fragment.

2.2.2 Radioactive Labelling of DNA for *DNase*I Footprinting

The 5' end of the I56i 436-187 fragment of DNA was filled using the large fragment of *E. coli* DNA polymerase I (Invitrogen). This was completed using 0.5 mM dNTP-dATP, 50μCi α-^{32}P dATP, 1X React2 (Invitrogen), 6U Polymerase and 100ng of DNA. The reaction was incubated at 37°C for one hour. The DNA was cleaned with 1 volume of phenol followed by 1 volume of chloroform. The DNA was precipitated in 5μg glycogen, 1/10 volume 3M NaOAc, and 2 volumes of EtOH for 1 hour at -20°C. The DNA was pelleted by centrifugation at 10,000g for 10 minutes, washed with 70% ethanol, dried and resuspended in 30μL ddH_2O. The labelled DNA was purified from a 1% low melting point agarose gel using β-agarase (New England Biolabs) following the manufacturer's guidelines.

2.2.3 *DNase*I Footprinting

Protein binding reactions (25μL protein buffer (20% glycerol, 4% polyvinyl alcohol, 1mM dithiothreitol, 20mM HEPES pH 7.9), 60mM KCl, 20μg/mL pdIdC, nuclear extracts and 1-2ng of labelled DNA) were performed in 50μL volumes and incubated on ice for 20 minutes. One volume of cofactor mix (10mM $MgCl_2$ and 5mM $CaCl_2$) was added followed by increasing amounts of *DNase*I, from 0.01 to 0.1 Kunitz units (Worthington) and incubated at room temperature for 2 minutes. The reaction was terminated by adding 1 volume of stop solution (1% SDS, 200mM NaCl, 20mM EDTA, 0.2mg/mL tRNA). Reactions were cleaned with 1 volume of phenol followed by 1 volume of chloroform and precipitated with 2 volumes of ethanol at -20°C for 1 hour. DNA was collected by centrifugation at 10,000g for 5 minutes, washed with 70% ethanol, and air dried. DNA was resuspended in 7μL formamide loading buffer (80% deionized formamide, 10mM EDTA, 1mg/mL xylene cyanoll, 1mg/mL bromophenol blue).

Reactions were loaded onto a 6% polyacrylamide sequencing gel (19/1% acrylamide/bis-acrylamide, 1XTBE [90 mM Tris base, 90mM Boric Acid, 2mM EDTA] and 7M urea). Guanine/adenine sequencing reactions were performed as previously described (Maxam and Gilbert 1980) and run on the same gel as a sequence marker. The gel was run at 80V in 1XTBE for 90 minutes and dried for 90 minutes. Radioactive bands were visualized by phosophorimaging using Imaging Screen-K and a Molecular Imager FX (BioRad).

Regions of probe protected by nuclear extracts were determined through comparison of DNA digestion patterns with and without nuclear extracts. The location of the protected region was determined by comparison to the GA sequencing reaction. MatInspector (Genomatix), a database of known transcription factor binding sites, was used to identify

putative binding sites within the protected regions. Protected regions were also compared to I56i sequence alignments (Ghanem, Jarinova et al. 2003) to identify areas of conservation within these regions.

2.3 Mutagenesis

For footprint sites FP1-FP5 and FP7, three to four base pair mistmatches were introduced within the core binding site identified. The areas covered by FP6 and FP8 are large and the interspecies conservation is extensive. Therefore, it was difficult to pinpoint only four nucleotides for mutagenesis and eight base pair mismatches were introduced to target multiple potential binding sites within the protected regions. With the exception of footprint site 8 (FP8), base-pair mutations were inserted into the putative binding sites by PCR of overlapping fragments. The first fragment began at the first base pair of the I56i enhancer and continued to a position approximately 10 base pair past the site of the mutation. The second fragment began at position 84 and continued to the end of the enhancer (base pair 400). *Xba*I restriction sites were incorporated into the mutations for screening purposes. Given the position of footprint site 8 at the end of the enhancer, a single PCR was completed encompassing the entire enhancer sequence. The base-pair mutations were incorporated into the primer sequence at the end of the enhancer. The primers used were 1.For, 84.For, Mut1.Rev, Mut2.Rev, Mut3.Rev, Mut4.Rev, Mut5.Rev, Mut6.Rev, Mut7.Rev, Mut8.Rev, and 400.Rev (Table 2.1). The conditions for the PCR are described in section 2.2.1 with annealing temperatures of 50-60°C.

PCR products were purified through gel extraction using the QIAquick Gel Extraction Kit (Qiagen) following manufacturer's protocols. The recovered DNA was used as template DNA for a subsequent PCR to create a full-length enhancer fragment containing

the mutant base-pairs. The primers used for this second PCR were 1.For and 400.Rev. The mutant fragments were ligated into the pDrive cloning vector (Qiagen) and transformed into XL-10 *E. coli* chemically competent cells (described in 2.2.1). Plasmid DNA was recovered and presence of the I56i enhancer containing the mutant binding site was verified by digestion with *Xba*I and *Eco*RI following manufacturer's protocols (Invitrogen).

2.4 Preparation of Reporter Constructs

Positive clones were grown in 2mL of LB overnight at 37°C with shaking and plasmid DNA was collected using the Wizard Plus© Plasmid Purification Kit (Promega). The enhancer was recovered through an *Eco*RI digestion and purified from agarose gel. The purified mutant enhancer was cloned into the p1230 vector (Yee and Rigby 1993) and again transformed into XL-10 *E. coli* chemically competent cells. The vector was prepared by an *Eco*RI partial digest to target the insertion of the enhancer into the correct location. This vector contains a *lacZ* reporter and a human *β-globin* minimal promoter. Plasmid DNA was collected and positive clones were identified by digestions with *Eco*RI, *Xba*I, and *Pst*I (Invitrogen).

Positive clones were regrown in 100mL of LB and plasmid DNA was extracted using the Qiagen Plasmid Midi Kit following the manufacturers' protocols. The mutant enhancer/*lacZ* reporter cassette was removed from the plasmid by digestion with *Sst*II and *Xho*I. The digestion was run on a 1% TAE (40 mM Tris Acetate, 1mM EDTA) gel and the fragment (~4kb) was recovered using GeneCapsule (GBiosciences). The fragment was precipitated with 0.3M NaCl and 2 volumes of 100% EtOH at -20°C for 30 minutes. The DNA was pelleted by centrifugation at 17,000g for 10 minutes, washed in 70% EtOH, and resuspended in 50μL TE (100mM Tris-Cl, 10mM EDTA). Dialysis was performed overnight

using a 0.025μm mixed cellulose ester membrane (Millipore) in TE at room temperature. DNA was collected in TE and diluted to 5ng/μL using PBS (137mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 2mM KH_2PO_4). Finally, the DNA was filtered in a 0.2μm cellulose acetate Centrex column (Whatman) at 17,000g for 5 minutes.

2.5 Production of Transgenic Mice

The mutant enhancer/*lacZ* reporter construct was sent for injection into CD1 mouse eggs at the Ottawa Health Research Institute. This was completed using standard procedures (Hogan, Costantini et al. 1986). Genomic DNA was prepared from the placenta (primary embryos) or from tail/ear tissue (transgenic lines) Tissue was digested at 55°C overnight in digestion buffer (50mM Tris-HCl pH8, 100mM EDTA, 100mM NaCl, 1%SDS) and 100-200ng Proteinase K. DNA was precipitated using ½ volume 5M NaCl followed by 0.8 volumes isopropanol. The DNA was pelletted by centrifugation at 10,000g for 2 minutes and washed with 70% EtOH. The DNA was air dried and resuspended in TE. Presence of the transgene was analyzed in primary embryos and transgenic lines via PCR. PCR was conducted using the genomic DNA as template. The primers used were Bg.For and lacZ.Rev as well as FH.X1 and FH.X2 as an internal control. PCR conditions were as described above (see 2.2.1) with an annealing temperature of 55°C.

Embryos for primary transgenics were harvested at E11.5. Embryos for transgenic lines were harvested at E10.5-E13.5. Embryos were fixed for 30 minutes at 4°C in fixing solution (1% formaldehyde, 0.2% gluteraldehyde, 0.02% Nonidet P-40 (NP-40), and 1X PBS). Embryos were then washed twice in PBS with agitation for 20 minutes. β-galactosidase staining was completed overnight in the dark at 30°C in staining solution (1mg/mL X-gal, 5mM $K_3Fe(CN)_6$, 5mM $K_4Fe(CN)_6$, 2mM $MgCl_2$, 0.02% NP-40 and

1XPBS). After staining, the embryos were washed with PBS and stored in 1XPBS and 10mM EDTA at 4°C. (Note: For E12.5 and E13.5 the brain was dissected prior to fixing for improved staining).

2.6 Sectioning

For each of the transgenic lines produced, E11.5 embryos were sectioned following β-galactosidase staining. The embryos were embedded in 1.5% agar and 5% sucrose and then soaked in 30% sucrose overnight at 4°C. Embryos were then mounted in Shandon Cryomatrix (Thermo Scientific) and stored at -80°C for at least 30 minutes. Sections were prepared at 30μm using a Leica CM1850. Sections were mounted on Superfrost/Plus slides (Fisher) using Aquatex (EMD Chemicals).

2.7 Electrophoretic Mobility Shift Assays (EMSA)

2.7.1 Preparation of DNA Fragments

Electrophoretic mobility shift assays were performed using fragments of DNA and double-stranded oligonucleotides. For the large fragments consisting of the entire enhancer sequence, the DNA already prepared for mutagenesis was used (WT, Mut1, Mut2, Mut3). The fragment was removed from the plasmid by *Eco*RI digestion and purified on agarose gel. The 150-300 base-pair overlapping fragments were prepared by PCR using the wild-type enhancer or the mutant enhancer (Mut1, Mut2, Mut3, Mut4) as template DNA. The primer combinations used were 24.For and 316.Rev (WT, Mut1, Mut2, Mut3), 187.For and 436.Rev (WT, Mut3, Mut4) and 187.For and 316.Rev (WT). The conditions for PCR were as described above (see 2.2.1), with an annealing temperature of 50°C. PCR products were purified from agarose gel, cloned into the pDrive vector and transformed into XL-10 *E. coli* chemically competent cells. Cells were grown overnight at 37°C with shaking in 100mL of

LB and plasmid DNA was recovered using the Qiagen Plasmid Midi Kit. The DNA fragment was removed from plasmid DNA by an *Eco*RI digestion and purified from an agarose gel.

Double-stranded oligonucleotides corresponding to selected areas of the identified footprint sites (see Table 2.1) were annealed via incubation at 95°C for 10 minutes followed by cooling to room-temperature overnight. They were then diluted to 50ng/μL in water.

2.7.2 Radioactive Labelling of DNA for EMSA

For the full enhancer and large fragment probes, the protruding 5' ends of the DNA were filled using the large fragment of *E. coli* DNA polymerase I (Invitrogen). This was completed using 0.5 mM dNTP-dATP, 50μCi α-^{32}P dATP, 1X React2 (Invitrogen), 6U Polymerase and 100-500ng of DNA. The reaction was incubated at 37°C for one hour. The DNA was cleaned with 1 volume of phenol followed by 1 volume of chloroform. The DNA was precipitated in 5μg glycogen, 1/10 volume 3M NaOAc, and 2 volumes of 100% EtOH for 1 hour at -20°C. The DNA was pelleted by centrifugation at 10,000g for 10 minutes, washed with 70% ethanol and reprecipitated in 1/10 volume 3M NaOAc and 2 volumes 100% EtOH for 1 hour at -20°C. The DNA was again pelleted by centrifugation at 10,000g for 10 minutes, washed with 70% EtOH, and resuspended in 50μL water.

50ng of double-stranded oligonucleotide probes were end-labelled with 50μCi γ -^{32}P dATP using 20 units of T4 polynucleotide kinase (New England Biolabs) and 1X kinase buffer. The DNA was cleaned with 1 volume of phenol followed by 1 volume of chloroform. The DNA was precipitated in 5μg tRNA, 1/10 volume 3M NaOAc, and 2 volumes of EtOH for 1 hour at -20°C. The DNA was pelleted by centrifugation at 10,000g for 10 minutes, washed with 70% ethanol and reprecipitated in 1/10 volume 3M NaOAc and 2 volumes

EtOH for 1 hour at -20°C. The DNA was again pelleted by centrifugation at 10,000g for 10 minutes, washed with 70% ethanol, and finally resuspended in 50μL water.

2.7.3 Electrophoretic Mobility Shift Assays

EMSAs were completed by incubating 1-2ng labelled DNA with the nuclear extracts prepared above in reaction buffer (7mM Tris-HCl pH 7.5, 80mM NaCl, 5mM $MgCl_2$, 10% glycerol, 0.05% NP-40, 1mg/mL bovine serum albumin, 2.75mM dithiothreitol) and 50μg/mL polydIdC. The incubation was performed at 0°C for 30 minutes. The reactions were loaded on a polyacrylamide gel (6% acrylamide/0.16% bis-acrylamide: oligonucleotide probes; 4%acrylamide/0.11% bis-acrylamide: full enhancer and large fragment probes) and run at 110V in 0.5X TBE for approximately 3 hours. The gel was dried and radioactive bands were visualized by phosophorimaging using Imaging Screen-K on a Molecular Imager FX (BioRad).

3. RESULTS

3.1 Nuclear Extracts from Mouse Brain/Forebrain Bind to DNA

Nuclear protein extracts were prepared from the brains of E11.5 mouse embryos and the forebrains of E13.5 mouse embryos. An electrophoretic mobility shift assay (EMSA) using an SP1 oligonucleotide probe was conducted in the absence and presence of the nuclear extracts to test their ability to bind DNA. Both extracts (E11.5 brain and E13.5 forebrain) created protein-DNA complexes of lower motility as seen by the presence of new bands on the gel (Figure 3.1).

3.2 *DNase*I Footprinting Identifies Putative Binding Sites in the I56i Enhancer

In the footprint experiments previously completed for the I56i enhancer sequence (see section 1.11), the protected region identified as FP6 was located close to one end of the sequence (Figure 1.8 a, b). Due to this location, the resolution of the footprint gels for the *DNase*I footprinting experiments already performed was not fine enough to identify the specific location of the putative binding site to the base pair level. Therefore, the *DNase*I footprinting was repeated on the same fragment of the I56i enhancer sequence; however, the DNA was labelled at the opposite end of the sequence to increase the resolution in the region of interest. Using nuclear extracts from the forebrain of E13.5 mouse embryos, the presence of 2-3 new protected regions (FP6, FP7, FP8) were identified (Figure 3.2). Two of the protected regions (FP6 and FP7) were consistently found in repeated experiments; however, the protected region FP8 could not be confirmed in repeated experiments. This protected region is located in a region of the probe that generates weakly labelled bands upon *DNase*I treatment, both in the presence and absence of nuclear proteins.

Figure 3.1. Electrophoretic mobility shift assay with a SP1 oligonucleotide probe and nuclear extracts prepared from E13.5 forebrain (a) and E11.5 brain (b). Arrows show bands of lower motility, demonstrating that proteins from both extracts bind DNA.

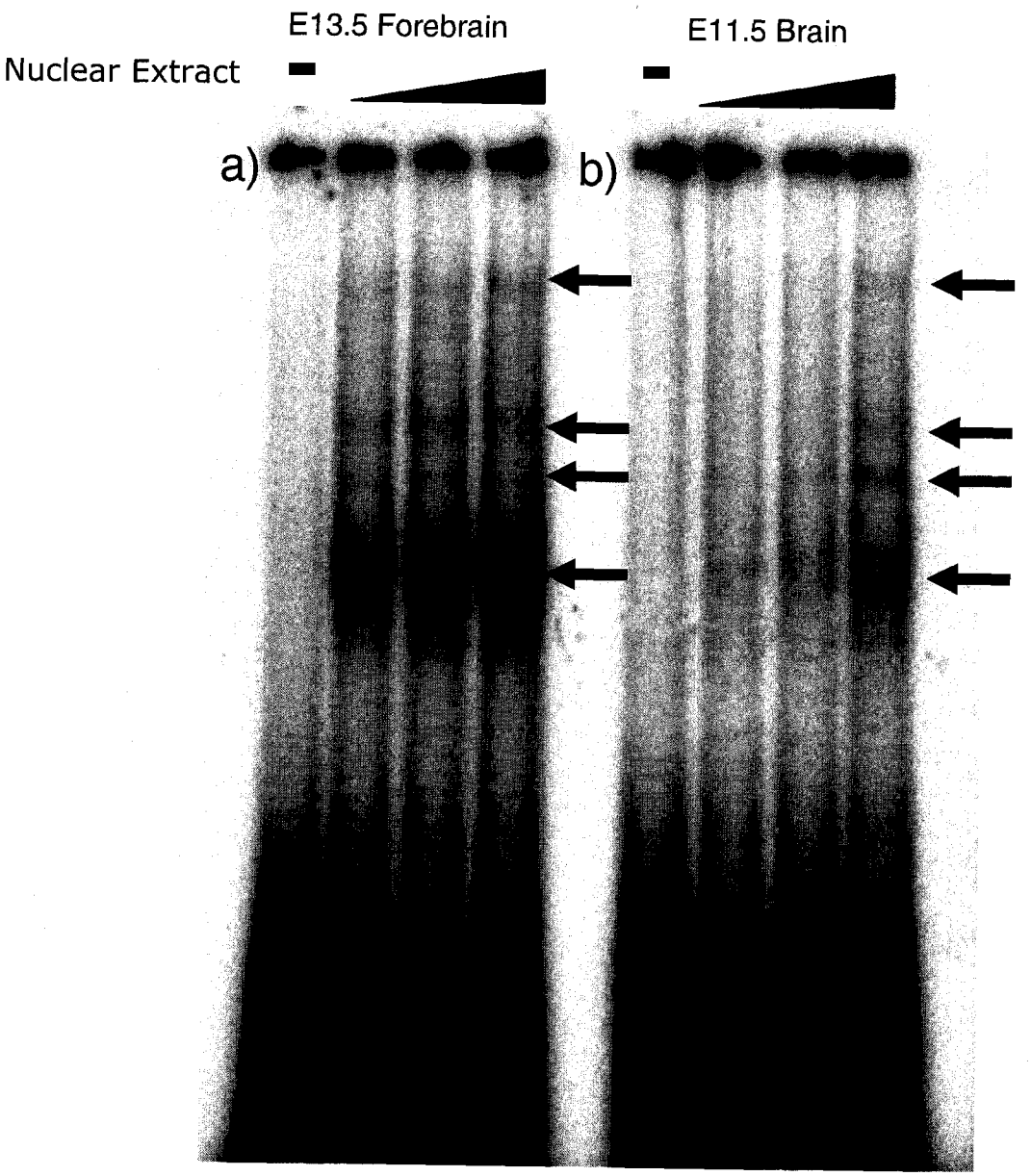
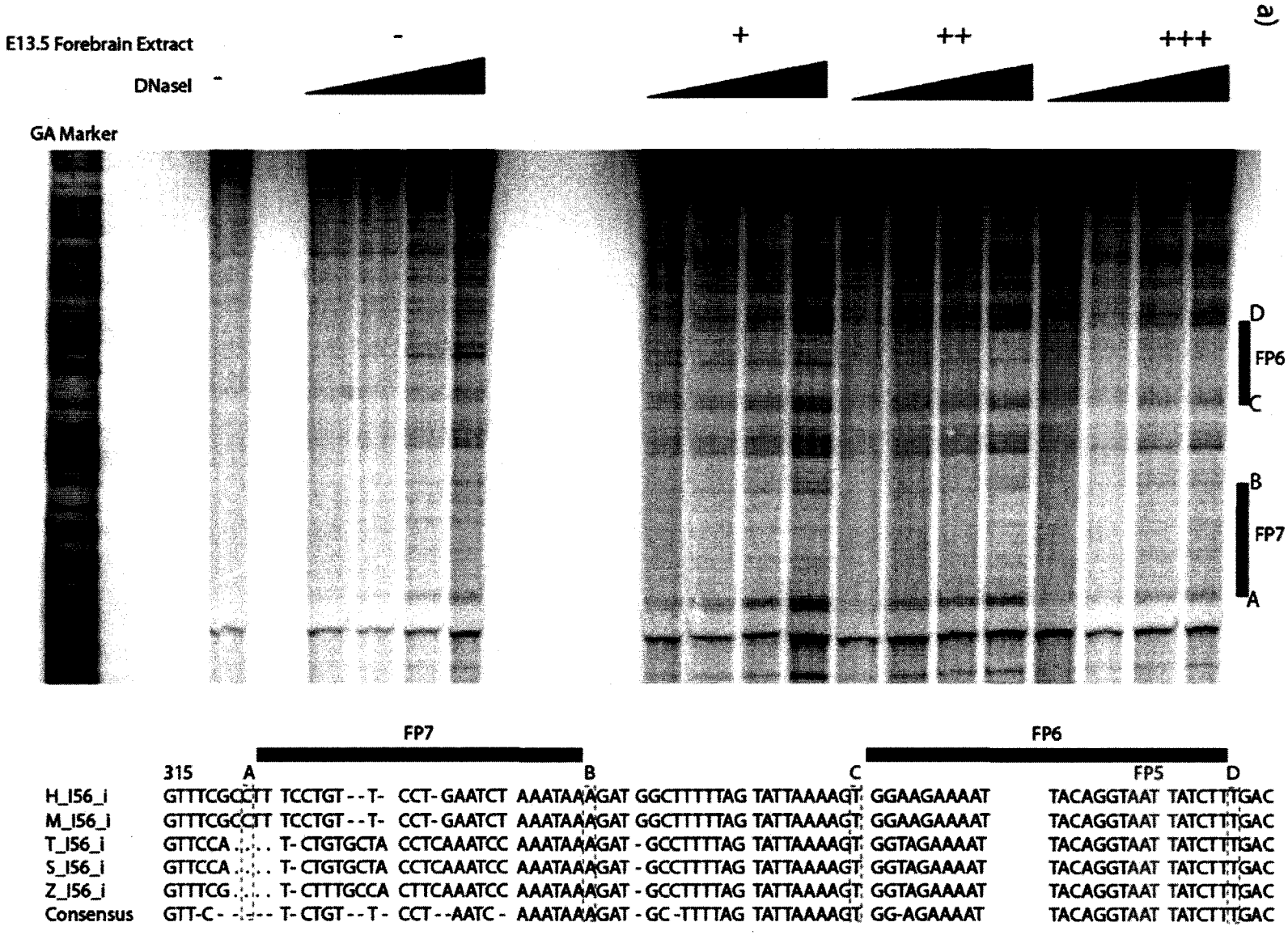
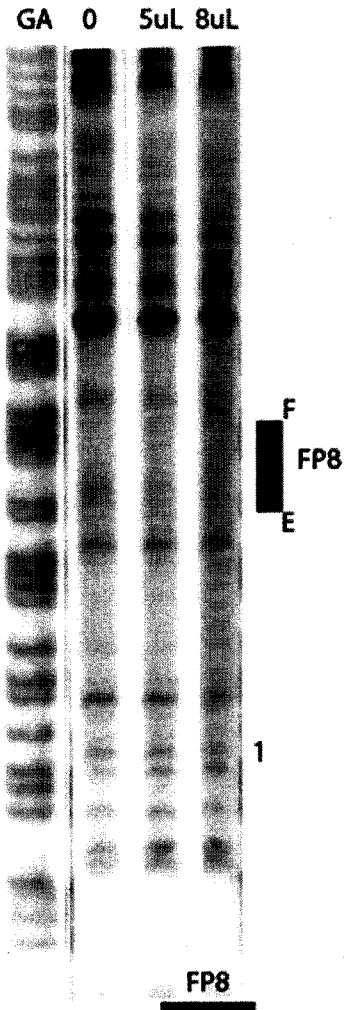

E13.5 Forebrain
E11.5 Brain
Nuclear Extract
a)
b)

Figure 3.2. *DNaseI* footprinting identifies three putative binding sites within the I56i enhancer. Proteins from a nuclear extract of E13.5 mouse forebrain bind one fragment of the mouse I56i enhancer sequence at three sites, FP6, FP7, and FP8. a) Identification of footprint sites FP6 and FP7. FP6 ranges from C to D and FP7 ranges from A to B. Increasing concentrations of nuclear extract are denoted by +, ++ and +++. Increasing concentrations of *DNase*I are denoted by the increasing slope. The previously identified TAAT core of the putative binding site within the FP5 protected region is shown in the sequence in blue. Data shown is representative of 3 repetitions. b) Identification of footprint site FP8. FP8 ranges from E to F. The volume of nuclear protein extracts used is given above the footprint. An equal concentration of *DNase*I was used in each lane. FP8 could not be reproduced in repeated experiments.

The sequencing GA marker is shown on the left of each footprint. Thick lines denote the protected sites. The location of the footprint sites within the I56i enhancer sequence is shown below the footprint in a sequence alignment for five species [H = human, M = mouse, T = *Takifugu rubripes*, S = *Spheroides nephelus*, Z = zebrafish (*Danio rerio*)].


a)
E13.5 Forebrain Extract
-
+
++
+++
DNaseI
-
GA Marker
D
FP6
C
B
FP7
A
FP7
FP6
315
A
B
C
FP5
D
H_I56_i GTTTCGCCTT TCCTGT--T- CCT-GAATCT AAATAAAGAT GGCTTTTTAG TATTAAAAGT GGAAGAAAAT TACAGGTAAT TATCTTTGAC
M_I56_i GTTTCGCCTT TCCTGT--T- CCT-GAATCT AAATAAAGAT GGCTTTTTAG TATTAAAAGT GGAAGAAAAT TACAGGTAAT TATCTTTGAC
T_I56_i GTTCCA..... T-CTGTGCTA CCTCAAATCC AAATAAAGAT -GCCTTTTAG TATTAAAAGT GGTAGAAAAT TACAGGTAAT TATCTTTGAC
S_I56_i GTTCCA..... T-CTGTGCTA CCTCAAATCC AAATAAAGAT -GCCTTTTAG TATTAAAAGT GGTAGAAAAT TACAGGTAAT TATCTTTGAC
Z_I56_i GTTTCG..... T-CTTTGCCA CTTCAAATCC AAATAAAGAT -GCCTTTTAG TATTAAAAGT GGTAGAAAAT TACAGGTAAT TATCTTTGAC
Consensus GTT-C- ----- T-CTGT--T- CCT--AATC- AAATAAAGAT -GC-TTTTAG TATTAAAAGT GG-AGAAAAT TACAGGTAAT TATCTTTGAC

b)



FP8

	1	E	F	50
H_I56_i	GCATTATAAT	TTTGGTTAAT	TTTCAATTTT	AG-G-TC-CT ACGTCTCT--
M_I56_i	GCATTATAAT	TTTGGTTAAT	TTTCAATTTT	AG-G-TC-CT ACGTCTCT--
T_I56_i	GCGTCGTAAT	TTCAGATAAT	ATCCCG----	AGCGCTCACT -C--CTCTCC
S_I56_i	GCGTCGTAAT	TTCAGATAAT	ATCCCG----	AGCGCTCACT -C--CTCTCC
Z_I56_i	ACATTGTAAT	TTTAGATAAT	ATCCCA----	AGCGTTCACT -C-TC-CT-C
Consensus	GC-T--TAAT	TT--G-TAAT	-T-C------	AG-G-TC-CT -C--CTCT--

*DNase*I footprinting was repeated using the same fragment that was used to identify the FP6, FP7, and FP8 protected regions, but with nuclear extracts from the brain of E11.5 mouse embryos. The protected regions identified in the footprints using E11.5 nuclear extracts were consistent with two of the protected sites isolated in the footprint experiments using the nuclear extracts prepared from E13.5 mouse forebrains (see Appendix 1). Protected regions were found in areas corresponding to the regions previously identified as FP6 and FP7. Within the region of the protected site FP8, identified above, the strength of the bands produced following *DNase*I treatment were again too weak to confirm the presence of this eighth protected region.

3.3 Identification of Putative Transcription Factor Binding Sites within Protected Regions

The sequence of the I56i enhancer was regularly compared to a library of known transcription factor-binding sites using the MatInspector software by Genomatix. Potential transcription factors binding within the protected regions located by *DNase*I footprinting were identified. Given the spatially restricted activity of the I56i enhancer, and to further limit the potential sites, the roles of the transcription factors identified were examined for expression in the developing forebrain. Transcription factors with a known expression domain in the brain or, preferentially, the forebrain, were chosen for further study ahead of transcription factors with no known expression in this region. A summary of the transcription factors identified within the protected sites and selected for further study is found in Table 3.1.

Previous research has already identified two DLX binding sites within the I56i enhancer sequence (Zerucha, Stuhmer et al. 2000). *DNase*I footprinting identified

Table 3.1. Summary of footprint sites and base pair changes introduced. The name of the footprint site (FP) and the corresponding mutagenesis site (Mut) are given. The sequence surrounding the targeted site is shown in both wild-type and mutant form. The position of the sequence within the 400 base pair enhancer sequence is given. If applicable, the transcription factor binding to the putative binding site is named. (Mut1-Mut4 constructs reprinted from Shipley 2005).

Footprint Site/ Mutant Construct	Sequence	Position Shown (bp)	Targeted Sites
FP1	TTT TAT CCA CAA	138-149	GATA-1
Mut1	TTT T cT agA CAA		
FP2	ATT CTC TTA AAT	164-175	Engrailed-1
Mut2	ATT CTa gaA AAT		
FP3	CAT AAT TAG AGT	181-192	DLX-1, -2, -5
Mut3	CAT c t a gAG AGT		
FP4	CGC TGA TTA CAG	209-220	Homeodomain
Mut4	CGC T c t a gA CAG		
FP5	AGA TAA TTA CCT	236-247	DLX-1, -2, -5
Mut5	AGA T c t a gA CCT		
FP6	TAC CTG TAA TTT TCT TCC ACT TTT AAT ACT AA	243-274	Conserved homeodomain
Mut6	TAC CT c TA g aTg TC g TCC AaT TTc tA g ACT AA		
FP7	ATT TAG ATT CAG GAA	287-301	Conserved
Mut7	ATT TA t cT a gAG GAA		
FP8	TGA AAA TTA ACC AAA ATT ATA ATG C	376-400	Conserved homeodomain
Mut8	TGA A t c T a g ACC AAA ATg g g g ATG C		

Note: Base pair mutations introduced are in red.

Position 1 of the I56i enhancer is defined as the base pair closest to the *Dlx6* gene.

protected regions (FP3, FP5) corresponding to the location of these putative DLX binding sites (Figure 1.8). Within the protected regions containing the putative DLX binding sites, binding sites for other homeodomain transcription factors were also identified. Within the sequence of the I56i enhancer containing the protected region FP3, putative binding sites were identified for GSH-1, GSH-2, HMX3/NKX5-1, NKX2.5/CSX, CART-1 and S8 type homeodomains. Within the sequence corresponding to the region of FP5, sites for PDX1, GATA-2, LIM/LHX3, and S8 type homeodomains were found.

Within the protected region of FP1, a sequence corresponding to GATA binding factor 1 was identified. A MEIS1b/HOXa9 heterodimer has also been shown to bind to a sequence found within the FP1 protected site in the I56i enhancer sequence.

A putative binding site for Engrailed-1 was identified within the FP2 protected region of the I56i enhancer sequence. Within the FP2 sequence, a putative binding site has also been identified for the HMX3/NKX5-1 homeodomain transcription factors.

The sequences of the protected regions identified in FP4, FP6, FP7 and FP8 did not initially map to specific transcription factors known to be expressed in the forebrain or other areas of endogenous *Dlx* expression. However, the sequences of three of the four protected sites (FP4, FP6 and FP8) contain the core homeodomain binding sequence, TAAT. The presence of these sequences is conserved between human, mouse, zebrafish and two species of pufferfish.

A potential binding site for HMX2/NKX5-2 was identified approximately 5 base pairs from the TAAT homeodomain core in FP4. Within the region of the I56i enhancer sequence identified as the FP6 protected region, a putative HMX3/NKX5-1 binding site was identified. This is also located three base pairs from the TAAT sequence previously

identified. A potential TST-1/OCT-6 binding site was also identified in the region of the FP6 protected site and the core of this binding site is identical to the TAAT core. No transcription factor binding sites were identified in the sequence of the I56i enhancer corresponding to the FP8 protected site.

The I56i enhancer sequence in the area of the protected region of FP7 does not contain a core homeodomain sequence nor were immediately relevant transcription factors initially identified. As a result, an area of conserved base pairs within the protected region was selected for further analysis. However, a putative binding site for BRN-2 has since been identified, just outside of the region previously selected for mutagenesis

3.4 Mutagenesis of Putative Binding Sites within the I56i Enhancer Sequence

Mutations were separately inserted into putative binding sites identified within each of the footprint sites using site-directed mutagenesis by PCR. Thus, for each of the eight protected regions (FP1-FP8) a mutant version of the I56i enhancer sequence containing targeted nucleotide changes within individual footprint sites was created and named Mut1-Mut8, respectively (see Table 3.1). A total of eight reporter constructs, one for each of the mutated enhancer sequences, were generated. (Mut1-Mut4 constructs reprinted from Shipley 2005).

3.5 Evaluation of the Effect of Binding Site Mutagenesis on I56i Enhancer Activity

Following mutagenesis of each of the eight putative binding sites, the 400 base pair I56i enhancer sequences containing one of the mutated footprint sites was introduced into a plasmid containing a β-globin minimal promoter-*lacZ* reporter cassette. This created a *lacZ* reporter construct for each of the eight mutagenized footprint sites.

The reporter constructs containing the mutagenized footprint sites 1-5 (Mut1, Mut2, Mut3, Mut4, Mut5) were injected into fertilized mouse eggs. Primary embryos were harvested at E11.5 or transgenic lines were generated by allowing the pregnancies to go to term. Pups were screened for the presence of the transgene and those carrying the transgene were grown to establish stable transgenic lines. When possible, adult mice carrying the transgene were mated and embryos were harvested at E10.5, E11.5, E12.5, and E13.5 in order to evaluate the effect of mutagenesis of the putative binding site on activity of the I56i enhancer throughout these developmental stages. A summary of the transgenic mice produced is shown in Table 3.2.

3.5.1 Activity of the Mutagenized I56i Enhancer in Whole Mount Embryos

The activity of the I56i enhancer with mutagenized putative binding sites was analyzed in whole mount embryos. Embryos obtained with the mutated enhancer construct were compared to those expressing a reporter construct carrying the wild-type I56i enhancer sequence. (wild-type construct from Ghanem, Jarinova et al. 2003).

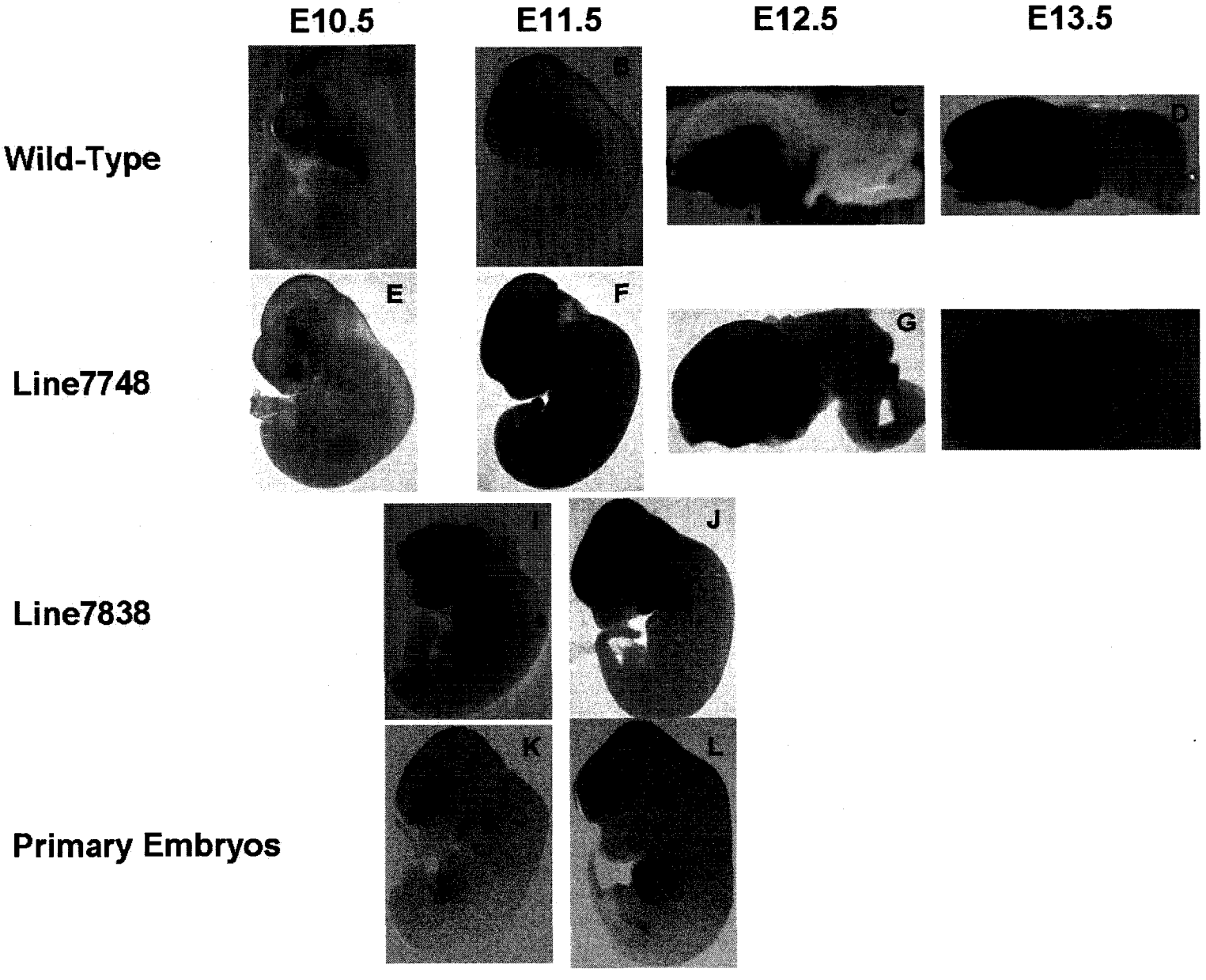
3.5.1a GATA-1 Binding Site (Mut1)

Four injections of the Mut1 construct resulted in two primary transgenic embryos and two stable lines expressing the transgene (7838 and 7748). The activity of the mutant enhancer construct in these embryos is shown in Figure 3.3. Embryos for line 7748 were harvested at E10.5, E11.5 E12.5, and E13.5 (Figure 3.3 E-H). The activity of the GATA-1 mutant enhancer appears to be similar to that of the wild-type enhancer in both the telencephalon and diencephalon. However, activity in the branchial arches appears to be drastically altered. β-galactosidase staining in the mandibular arch is almost entirely

Table 3.2. Summary of constructs injected and the number of transgenic primary embryos and transgenic lines produced. The name of the footprint site (FP) and the corresponding mutagenesis site (Mut) are given. Transgenic animals for the FP5-FP8 mutant constructs were not produced due to technical reasons, discussed in section 4.1.5a.

Footprint Site	Mutant Construct	# of Injections	# Primary Embryos	# Transgenic Lines
FP1	Mut1	4	2	2
FP2	Mut2	4	1	2
FP3	Mut3	4	3	0
FP4	Mut4	6	6	1
FP5	Mut5	1	n/a	0
FP6	Mut6	0	not injected	
FP7	Mut7	0	not injected	
FP8	Mut8	0	not injected	

Figure 3.3. Activity of the transgenic mice expressing the Mut1 I56i *lacZ* reporter construct in comparison to mice expressing the wild-type I56i construct. Embryos are shown at E10.5 and E11.5 and dissected brains are shown at E12.5 and E13.5. Embryos carrying the wild-type I56i construct show activity in the forebrain (A-D) and the mandibular and hyoid arches (A-B). Both transgenic lines (7748 and 7838) carrying the Mut1 construct show activity in the forebrain similar to that seen in the wild-type constructs (E-J). However, both of the primary embryos (K,L) show weaker activity in both the telencephalon and diencephalon when compared to the wild-type activity (B). Activity in the branchial arches in three of the four Mut1 transgenic mice is reduced or abolished (E, F, K, L).

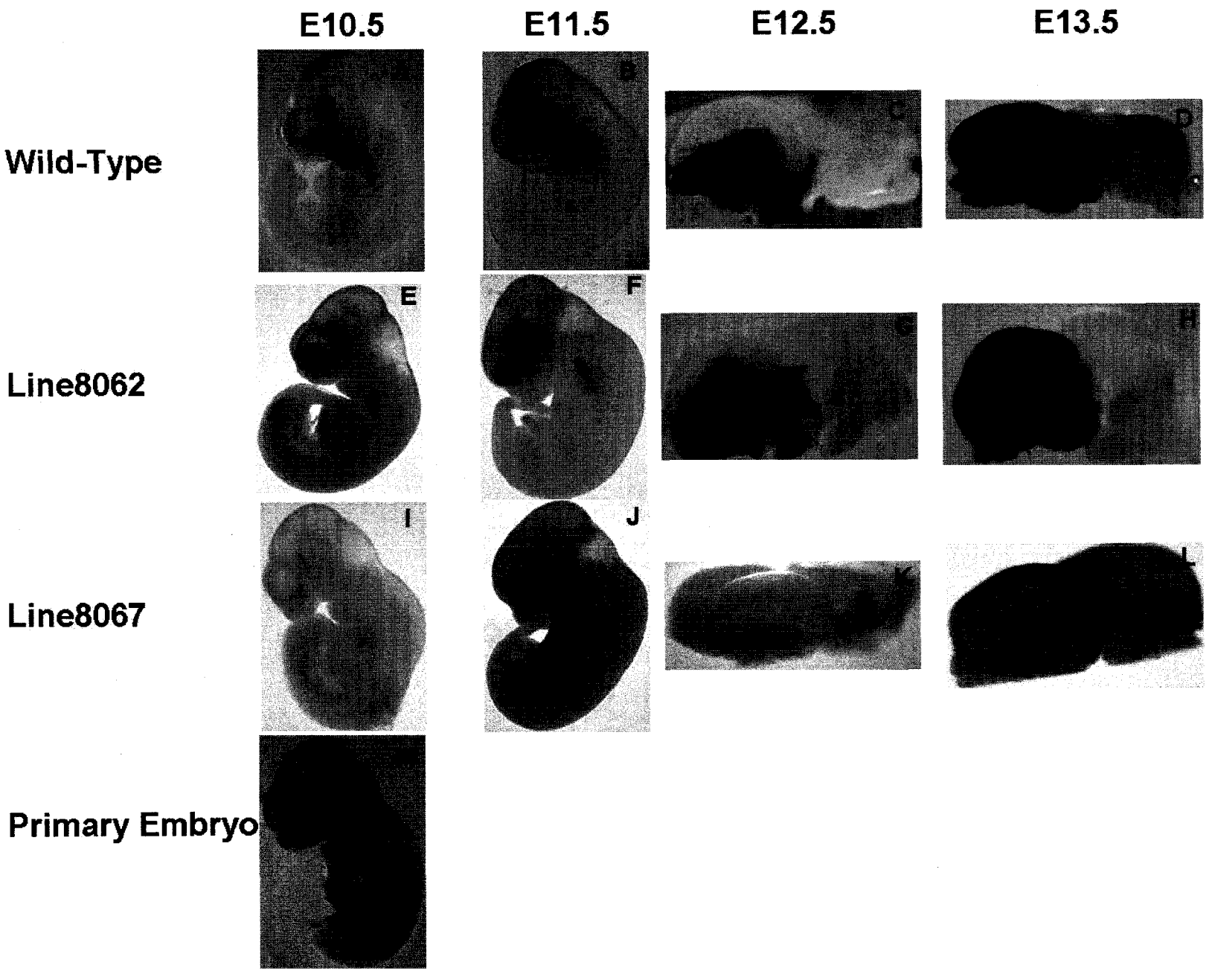

E10.5
E11.5
E12.5
E13.5
Wild-Type
Line7748
Line7838
Primary Embryos
A
B
C
D
E
F
G
I
J
K
L

eliminated and staining in the hyoid arch is greatly reduced when compared to that seen in the wild-type transgenic embryos. In contrast, the activity seen in line 7838 (Figure 3.3 I, J), which was only evaluated at E11.5, is very similar to the activity of the wild-type enhancer. Both areas of the forebrain show activity similar to that in the wild-type and the effect on activity in the branchial arches is not as dramatic as seen in line 7748. In the embryos of line 7838, the staining in the hyoid arch appears relatively normal and the mandibular arch shows only a slight reduction in staining. It should be noted that some embryos of line 7838 also show extra β-galactosidase staining in some areas of the head outside of the telencephalon and diencephalon. Finally, both primary transgenic embryos at E11.5 show similar alterations in the activity of the I56i enhancer (Figure 3.3 K, L). Activity in the telencephalon appears to be localized correctly but slightly reduced; however, the activity of the diencephalon appears to be greatly reduced. Areas corresponding to the mandibular arch and the hyoid arch show reductions in the activity of the enhancer in both primary transgenic embryos, with activity in the mandibular arch almost completely lost in one of the two embryos (Figure 3.3 L).

3.5.1b Engrailed-1 Binding Site (Mut2)

The construct containing the I56i enhancer with mutagenesis of the putative Engrailed-1 binding site was injected four times and resulted in one primary transgenic embryo and two stable transgenic lines (8062 and 8067) (Figure 3.4 E-M). Activity of the enhancer in the branchial arches is significantly decreased in the primary transgenic embryo and in embryos obtained from both transgenic lines. Activity in the mandibular arch is lost in all three and β-galactosidase staining in the hyoid arch is lost in two and reduced in the third. However, the effects of the putative binding site mutation in the forebrain are less consistent. Line 8062 shows activity in both the telencephalon and diencephalon very similar to that

Figure 3.4. Activity of the transgenic mice expressing the Mut2 I56i *lacZ* reporter construct in comparison to mice expressing the wild-type I56i construct. Embryos are shown at E10.5 and E11.5 and dissected brains are shown at E12.5 and E13.5. Embryos carrying the wild-type I56i construct show activity in the forebrain (A-D) and the mandibular and hyoid arches (A-B). The primary embryo (M) and line8067 (2050) carrying the Mut2 construct (I-L) show significant reductions in activity in the telencephalon and diencephalon. Both transgenic lines and the primary embryo carrying the Mut2 construct show a reduction or complete loss of activity in the branchial arches (E,F, I, J, M).


E10.5
E11.5
E12.5
E13.5
Wild-Type
Line8062
Line8067
Primary Embryo
A
B
C
D
E
F
G
H
I
J
K
L

seen in the reporter construct with the wild-type enhancer (Figure 3.4 E-H). On the other hand, line 8067 (Figure 3.4 I-L) and the primary transgenic embryo (Figure 3.4 M) both show a complete loss of reporter activity in the diencephalon and a drastic reduction of reporter activity in the telencephalon.

3.5.1c DLX Binding Site (Mut3)

Three primary embryos at E11.5 were produced following four injections of the reporter construct containing the mutation of one of the putative DLX binding sites within the I56i enhancer sequence (Figure 3.5). The effect of the mutation on enhancer activity is variable amongst the three primaries. In the telencephalon, one primary transgenic embryo (Figure 3.5 B) shows activity similar to that seen with the wild-type enhancer (Figure 3.5 A). A second primary transgenic embryo (Figure 3.5 C) shows a moderate reduction in activity and the third (Figure 3.5 D) shows a significant reduction in the telencephalon. In the diencephalon, two of the three primary transgenic embryos (Figure 3.5 B, C) show little or no change in activity when compared to the wild-type. The third (Figure 3.5 D), which showed the largest reduction in activity in the telencephalon, shows a complete loss of activity in the diencephalon. In the branchial arches, all three primary transgenic embryos lack activity in the mandibular arch. In addition, two of the three primary embryos show a complete loss of activity in the hyoid arch and one shows a moderate reduction in this area.

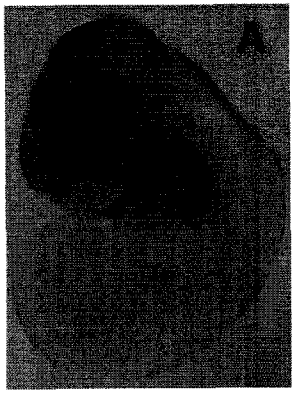
3.5.1d Putative Homeodomain Binding Site (Mut4)

The reporter construct with the enhancer sequence containing the mutation of the putative homeodomain binding site was injected six times and resulted in one transgenic line and 6 primary transgenic embryos (Figure 3.6). All primary embryos showed significant

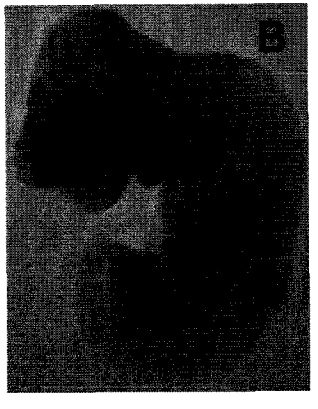
Figure 3.5. Activity of the transgenic mice expressing the Mut3 I56i *lacZ* reporter construct in comparison to mice expressing the wild-type I56i construct. Embryos are shown at E11.5. Embryos carrying the wild-type I56i construct show activity in the forebrain and the mandibular and hyoid arches (A). Two of the three primary transgenic embryos carrying the Mut3 construct show a reduction in activity in the forebrain (C, D). All three primaries also show a reduction or abolition of activity in the branchial arches (B, C, D).

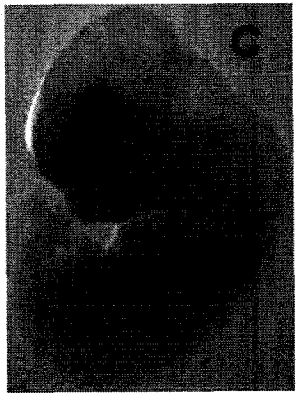
E11.5

Wild-Type



Primary Embryos





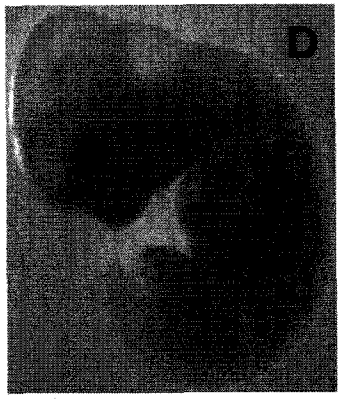
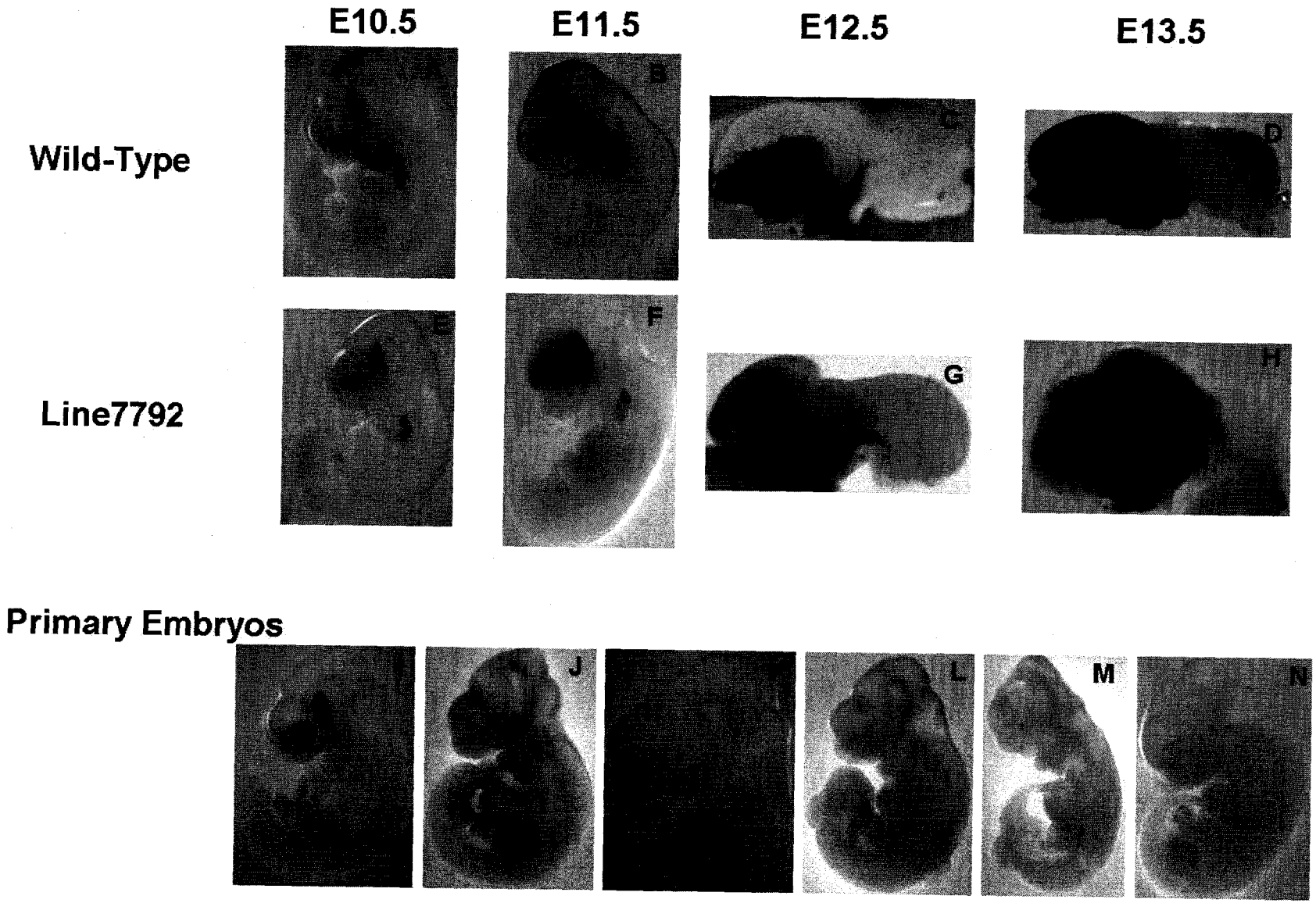


Figure 3.6. Activity of the transgenic mice expressing the Mut4 I56i *lacZ* reporter construct in comparison to mice expressing the wild-type I56i construct. Embryos are shown at E10.5 and E11.5 and dissected brains are shown at E12.5 and E13.5. Embryos carrying the wild-type I56i construct show activity in the forebrain (A-D) and the mandibular and hyoid arches (A-B). The transgenic line carrying the Mut4 construct shows activity in the forebrain comparable to that seen in the wild-type embryos (E-H) but a reduction in activity in the branchial arches (E, F). Of the primary transgenic embryos carrying the Mut4 construct, five of the six show a reduction in activity in the forebrain (K-N) and activity in the branchial arches is lost in all six (I-N) (Note: I-M at E10.5, N at E11.5).


E10.5
E11.5
E12.5
E13.5
Wild-Type
Line7792
A
B
C
D
E
F
G
H
Primary Embryos
I
J
K
L
M
N

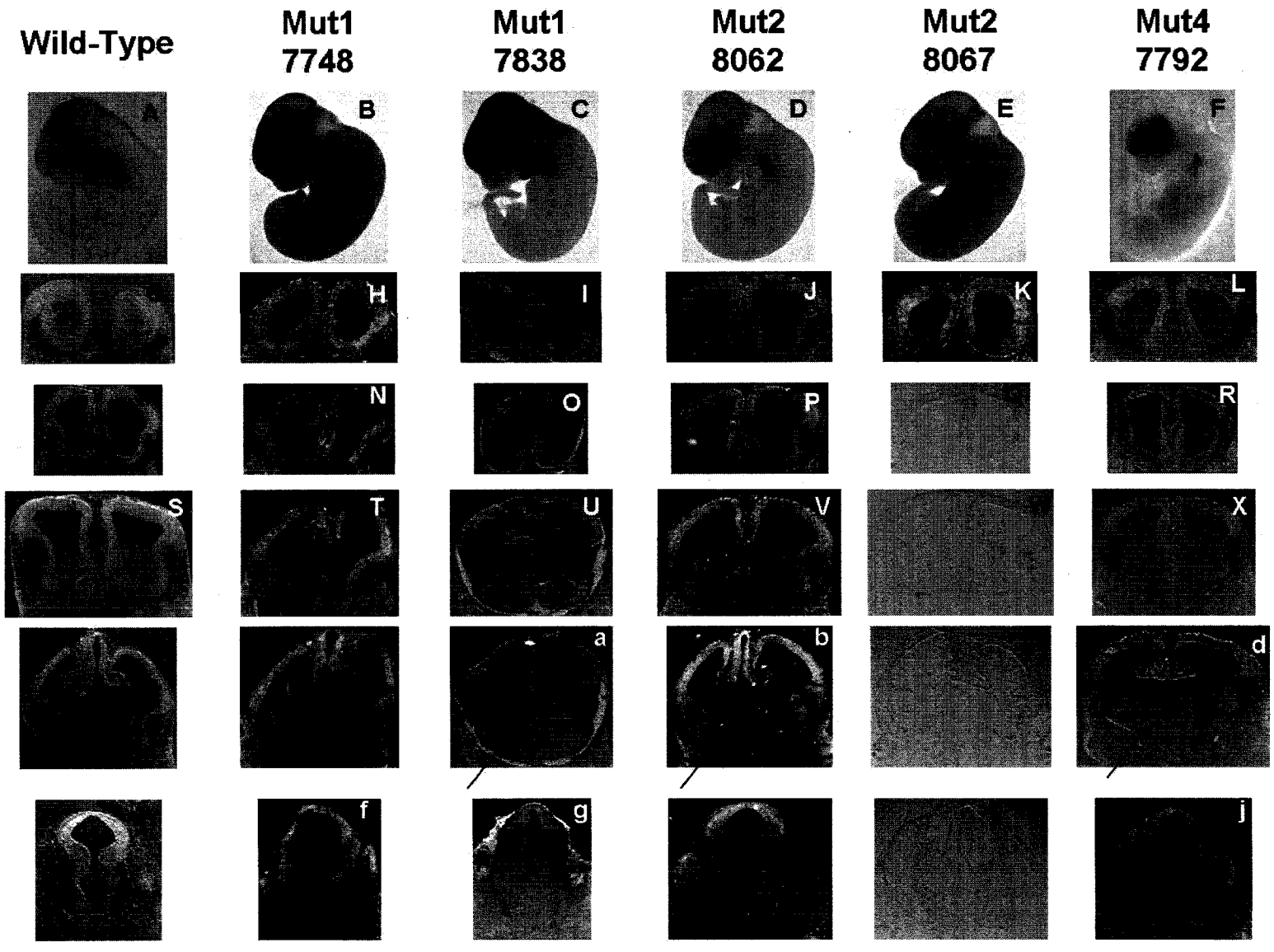
variation in activity when compared to the activity of the construct carrying the wild-type enhancer sequence. Activity in the telencephalon was greatly reduced in all of the primary embryos (Figure 3.6 I-N). Activity in the diencephalon was also lost in all but one of the primary embryos (Figure 3.6 I, K-N). In this last primary embryo (Figure 3.6 J) activity was only slightly reduced. β-galactosidase staining was also lost in the mandibular and hyoid arches in all the primary embryos examined. With respect to the forebrain, the transgenic line (7792) did not show results similar to that of the primary transgenic embryos. Activity in both the telencephalon and diencephalon appears normal in embryos harvested from line 7792 (Figure 3.6 E-H). However, like the primary transgenic embryos, the activity in the branchial arches in line 7792 is significantly affected. β-galactosidase staining in the mandibular arch is completely lost and the area of staining in the hyoid arch is greatly reduced (Figure 3.6 E, F).

3.5.2 Mutant Enhancer Activity in the Forebrain

Whole-mount embryos are useful to have a general view of enhancer activity in the forebrain but tissue sectioning is necessary to examine enhancer activity at the level of specific cells. To further evaluate if the activity of the enhancer was altered at this level following mutagenesis of the putative binding sites within the enhancer, E11.5 embryos of the transgenic lines produced were sectioned (Figure 3.7). At E11.5, activity of the wild-type I56i enhancer is seen in regions of the sub-ventricular zone and mantle zone within the medial ganglionic eminence (MGE) and the lateral ganglionic eminence (LGE) (Figure 3.7 M, S).

As with the whole-mount embryos, line 8067, carrying the construct with mutagenesis of the putative Engrailed-1 binding site (Mut2) appears to show the greatest

Figure 3.7. Sections of transgenic embryos carrying the wild-type and mutant I56i *lacZ* reporter constructs. 30μm sections are shown in an increasingly dorsal position from top to bottom. A whole-mount embryo is shown at the top as a representative for the sections. Activity of the wild-type enhancer is shown in the first column for comparison (A, G, M, S, Y, e). All sections from transgenic line 8067 carrying the Mut2 construct show a significant reduction in staining throughout all sections (K, Q, W, c, i). In addition, a reduction in staining is seen in the sub-ventricular zone (SVZ) of three of the lines (arrows in a, b, d).


Wild-Type
Mut1
7748
Mut1
7838
Mut2
8062
Mut2
8067
Mut4
7792
A
B
C
D
E
F
H
I
J
K
L
N
O
P
R
S
T
U
V
X
a
b
d
f
g
j

variation in forebrain activity when compared to the wild-type activity of the I56i enhancer. Unlike the wild-type transgenic embryos, which show large areas of β-galactosidase staining in the forebrain, each section from embryos of line 8067 shows only a few cells of reporter activity (Figure 3.7 E, K, Q, w, c, i). This apparent decrease in activity is consistent throughout the brain. In contrast, the other transgenic line established carrying the enhancer sequence with the Mut2 base pair changes (line 8062) shows a pattern of activity very similar to that seen with the wild-type enhancer sequence (Figure 3.7 D, J, P, V, b, h). A slight difference exists in that there appears to be a decrease of activity within the MGE (Figure 3.7 b), which shows a much fainter and less evenly spread activity in the mutant transgenic (line 8062) when compared to the wild-type transgenic line.

A similar reduction in activity in the MGE is seen in one of the two transgenic lines with mutagenesis of the putative GATA-1 binding (Figure 3.7 a). Line 7838 shows a slight decrease in the activity and a variation in the localization of β-galactosidase staining in the MGE and also shows a small decrease in activity in the LGE. However, the second transgenic line with the Mut1 construct (line 7748) shows β-galactosidase staining very similar to that seen in the line carrying the wild-type enhancer construct (Figure 3.7 B, H, N, T, Z, f).

Only one transgenic line was successfully obtained carrying the Mut4 construct (line 7792). In most areas of the forebrain, the activity of mutant enhancer construct is similar to that seen in the wild-type construct (Figure 3.7 F, L, R, X, j). A difference is again seen in the MGE which shows much fainter β-galactosidase staining in comparison to the staining seen in the wild-type transgenic embryos (Figure 3.7 d). Furthermore, faint activity appears in regions where the wild-type enhancer construct is not expressed (Figure 3.7 X, d).

3.6 *In vitro* Analysis of Mutagenesis of the Putative Binding Sites in the I56i Enhancer

3.6.1 150-400 Base Pair Probe EMSA

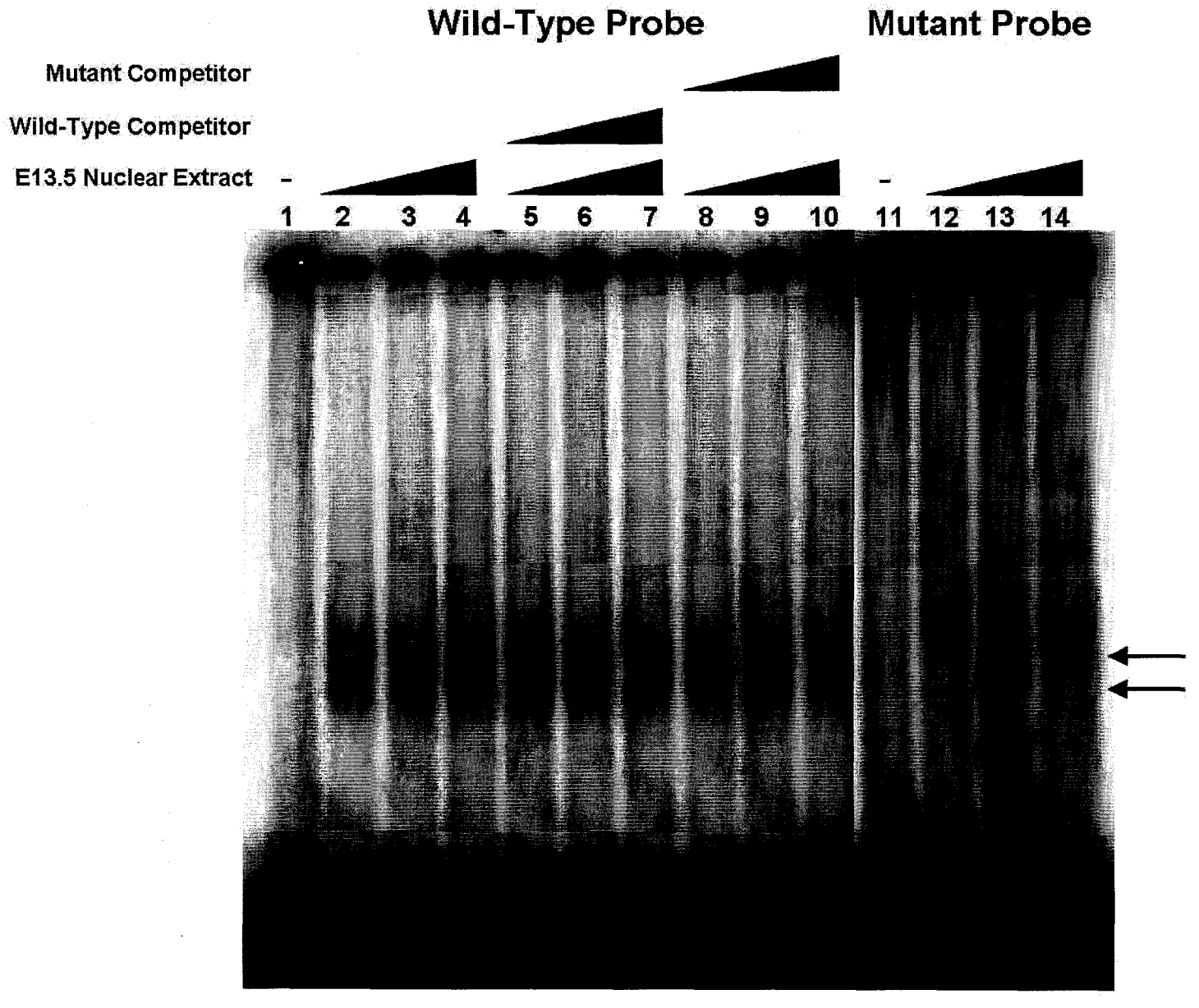
EMSA experiments were attempted using probes encompassing the entire 400 base pair enhancer sequence, as well as probes composed of 150-300 base pair fragments of the enhancer sequence. For each of the two types of probes, the ability of proteins to bind the wild-type sequence was compared to the ability of proteins to bind the mutagenized FP1-FP4 binding sites. For all 150-400 base pair probes, clear protein-DNA complexes could not be obtained. As a result, it was not possible to determine the effect mutagenesis of the putative protein binding sites using these probes. Although the reasons for this are still unclear, potential explanations are discussed below.

3.6.2 EMSA With Oligonucleotide Probes

EMSA experiments were repeated using 26 base pair oligonucleotide probes in an effort to improve the protein-DNA complexes and the quality of the results. This was completed for the first four protected regions (FP1-FP4). The oligonucleotide probes were designed to encompass the regions of the putative binding sites, with sequences corresponding to either the wild-type putative binding site or the mutated binding site sequence (a list of the oligonucleotides used to generate these probes is available in Table 2.1). Concurrently, competition assays were completed using both wild-type and mutant competitors.

In EMSA experiments using probes corresponding to protected region FP1, the wild-type probe shows the formation of much stronger DNA-protein complexes when compared to the probe carrying the mutated GATA-1 putative binding site (Mut1) (Figure 3.8 lanes 2-4 compared to lanes 12-14). This suggests that the proteins within the nuclear extract are not

Figure 3.8. Mutagenesis of the putative GATA-1 binding site appears to reduce the ability of E13.5 nuclear extracts from the forebrain to bind a fragment of the I56i enhancer. An electrophoretic mobility shift assay with a radioactively labelled oligonucleotide fragment of the I56i enhancer corresponding to an area of the protected region FP1 in wild-type form (lanes 1-10) or with the base-pair mutations corresponding to Mut1 (lanes 11-14). The intensity of the lower motility complexes (arrows) with the Mut1 probe (lanes 12-14) is lower than that seen using the wild-type probe (lanes 2-4). Competition using unlabelled wild-type oligonucleotide does not show as strong a difference compared to competition with unlabelled Mut1 oligonucleotide. The wild-type competitor (lane 7) shows only a slightly better ability to bind with the nuclear extracts when compared to the Mut1 competitor (lane 10). Increasing concentrations of nuclear extracts and unlabelled competitors are shown by increasing slopes. Lanes 5 and 8: 1:1 probe:competitor ratio; lanes 6 and 9: 10x molar excess competitor; lanes 7 and 10: 150x molar excess competitor. Data shown is representative of three repetitions.
Oligonucleotide sequences used: 1.19WT.For: 5'-TTAAGGTTTTTATCCACAAGAAGGT-T-3'; 1.19WT.Rev: 5'-AACCTTCTTGTGGATAAAAACCTTAA-3'; 1.19Mut.For: 5'-TTA-AGGTTTTTCTAGACAAGAAGGTT-3'; 1.19Mut.Rev: 5'-AACCTTCTTGTCTAGAAAA-ACCTTAA-3'


Wild-Type Probe
Mutant Probe
Mutant Competitor
Wild-Type Competitor
E13.5 Nuclear Extract
-
-
1 2 3 4 5 6 7 8 9 10 11 12 13 14

able to bind to the probe with the mutated putative binding site as well as they can bind to the probe with the wild-type sequence. The competition assay shows consistent, but less definitive, results. Competition with the unlabelled wild-type oligonucleotide reduces the strength of the band representing the complex of lower motility (Figure 3.8 lanes 5-7). In comparison, competition with the unlabelled mutant oligonucleotide also reduces the intensity of the band but to a slightly lesser degree (Figure 3.8 lanes 8-10). This suggests that mutagenesis of the putative binding site reduces but does not completely eliminate the ability of the proteins to bind this region of the I56i enhancer sequence.

In comparison to other sites that were mutated, mutagenesis of the putative Engrailed-1 binding site shows the greatest effect on protein binding. A corresponding effect on the ability of the mutant oligonucleotide to act as a competitor is also seen (Figure 3.9). The DNA-protein complexes formed with the mutant probe (Figure 3.9 lanes 12-14) are much weaker than those formed using the probe with the wild-type sequence (Figure 3.9 lanes 2-4). The wild-type competitor is also able to bind with the proteins and eliminate the visible complex of lower motility at a much lower concentration than the competitor with the mutated putative binding site (Figure 3.9 lanes 5-7 compared to lanes 8-10). Thus, it appears that the alteration of the putative Engrailed-1 binding site significantly reduces but does not eliminate the ability of proteins to bind to the 26 base pair area of the enhancer sequence in the region of the FP2 protected site.

The effect of mutagenesis of the putative binding sites in the protected regions FP3 (Figure 3.10) and FP4 (Figure 3.11) is not as pronounced as that described following mutagenesis of the Engrailed-1 putative binding site, described above. Mutagenesis of the putative DLX binding site in the protected region of FP3 reduces the ability of proteins to

bind the unlabelled competitor in only one of the two complexes of lower motility formed (Figure 3.10 lane 8). A moderate reduction is seen in the ability of proteins to bind a probe corresponding to the mutation of the putative homeodomain transcription factor binding site in the FP4 protected region (Figure 3.11 lanes 2-4 compared to lanes 12-14); however, the effect of the mutant competitor is not as evident as the reduction in the visible protein-DNA complex is similar to that seen with the wild-type competitor (Figure 3.11 lanes 5-7 compared to lanes 8-10).

Figure 3.9. Mutagenesis of the putative Engrailed-1 binding site appears to reduce the ability of E13.5 nuclear extracts from the forebrain to bind a fragment of the I56i enhancer. An electrophoretic mobility shift assay with a radioactively labelled oligonucleotide fragment of the I56i enhancer corresponding to part of the FP2 protected region in wild-type form (lanes 1-10) or with the base-pair mutations corresponding to Mut2 (lanes 11-14). The intensity of the DNA-protein complex (arrow) with the Mut2 probe (lanes 12-14) is lower than that seen using the wild-type probe (lanes 2-4). Competition with unlabelled oligonucleotides shows a greater ability of the proteins to bind the wild-type sequence in comparison to the Mut2 sequence as the reduction in the complex of lower motility is greater with the wild-type competitor (lanes 5-7) in comparison to the Mut2 competitor (lanes 8-10). Increasing concentrations of nuclear extracts and unlabelled competitors are shown by increasing slopes. Lanes 5 and 8: 1:1 probe:competitor ratio; lanes 6 and 9: 10x molar excess competitor; lanes 7 and 10: 100x molar excess competitor. Data shown is representative of five repetitions.

Oligonucleotide sequences used: 2.7WT.For: 5'-TTTTTCCATTCTCTTAAATGCAGCCA-3'; 2.7WT.Rev: 5'-TGGCTGCATTTAAGAGAATGGAAAAA-3'; 2.7Mut.For: 5'-TTTTT-CCATTCTAGAAAATGCCAGCCA-3'; 2.7Mut.Rev: 5'-TGGCTGCATTTTCTAGAATG-GAAAAA-3'

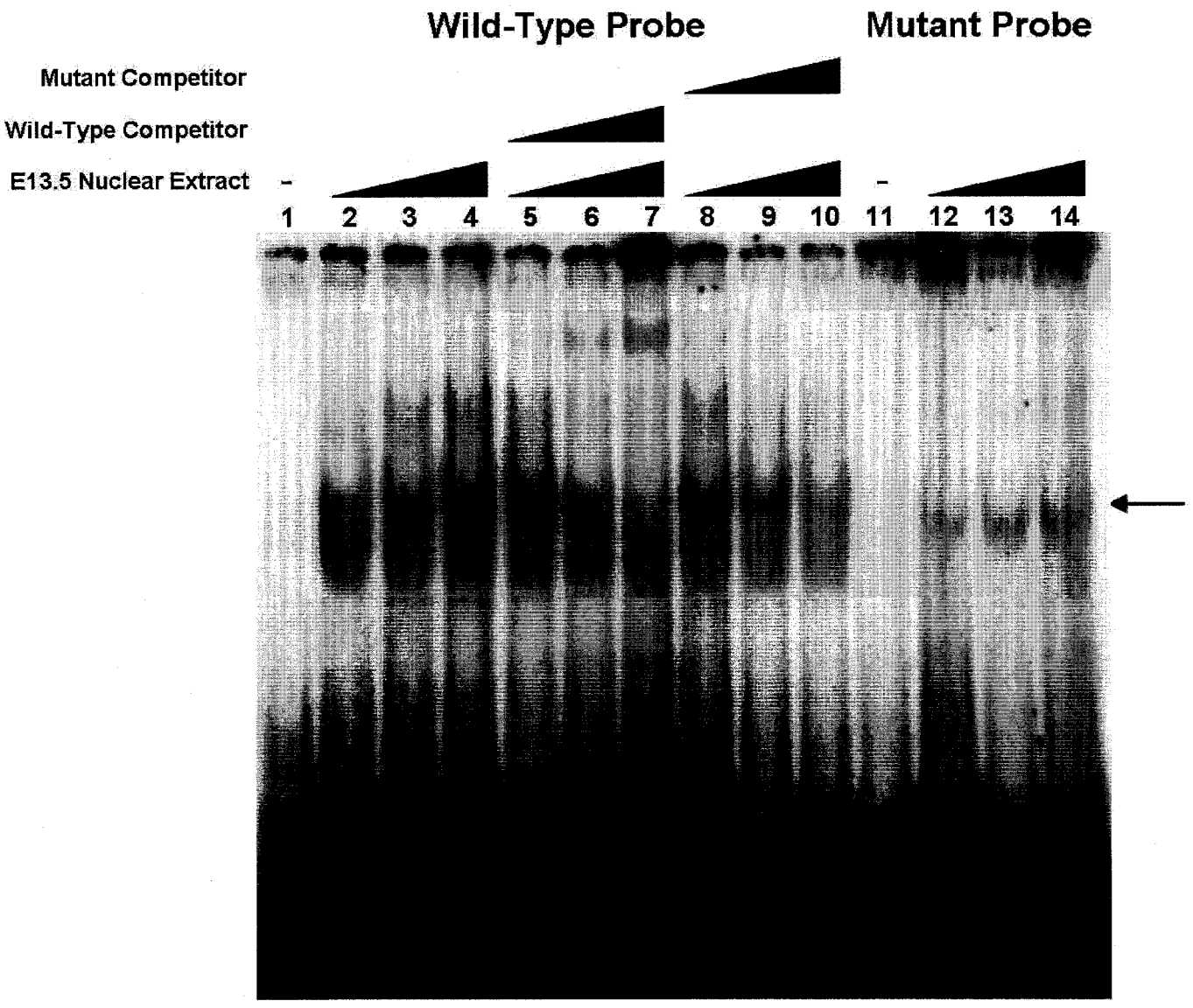

Wild-Type Probe
Mutant Probe
Mutant Competitor
Wild-Type Competitor
E13.5 Nuclear Extract
-
-
1 2 3 4 5 6 7 8 9 10 11 12 13 14

Figure 3.10. Mutagenesis of the putative DLX binding site in the FP3 protected region reduces the ability of proteins to bind a fragment of the I56i enhancer sequence. An electrophoretic mobility shift assay with a radioactively labelled oligonucleotide fragment of the I56i enhancer corresponding to a portion of the FP3 protected region. Competition with unlabelled oligonucleotides (both wild-type and Mut3) prevent binding of one of the two complexes of lower motility (lanes 3-8, top arrow). However, the reduction in the strength of the second complex of lower motility (lower arrow) is much greater with a wild-type competitor (lanes 3-5) when compared to the Mut3 competitor (lanes 6-8). Concentration of the competitors is denoted by the slopes. Presence/absence of nuclear extracts is shown by +/-. Lanes 3 and 6: 1:1 probe:competitor ratio; lanes 4 and 7: 10x molar excess competitor; lanes 5 and 8: 100x molar excess competitor. Data shown is representative of three repetitions.

Oligonucleotide sequences used: 3.9WT.For: 5'-AATGCAGCCATAATTAGAGTAATTTT-3'; 3.9WT.Rev: 5'-AAAATTACTCTAATTATGGCTGCATT-3'; 3.9Mut.For: 5'-AATGC-AGCCATCTAGAGAGTAATTTT-3'; 3.9Mut.Rev: 5'-AAAATTACTCTCTAGATGGCT-GCATT-3'

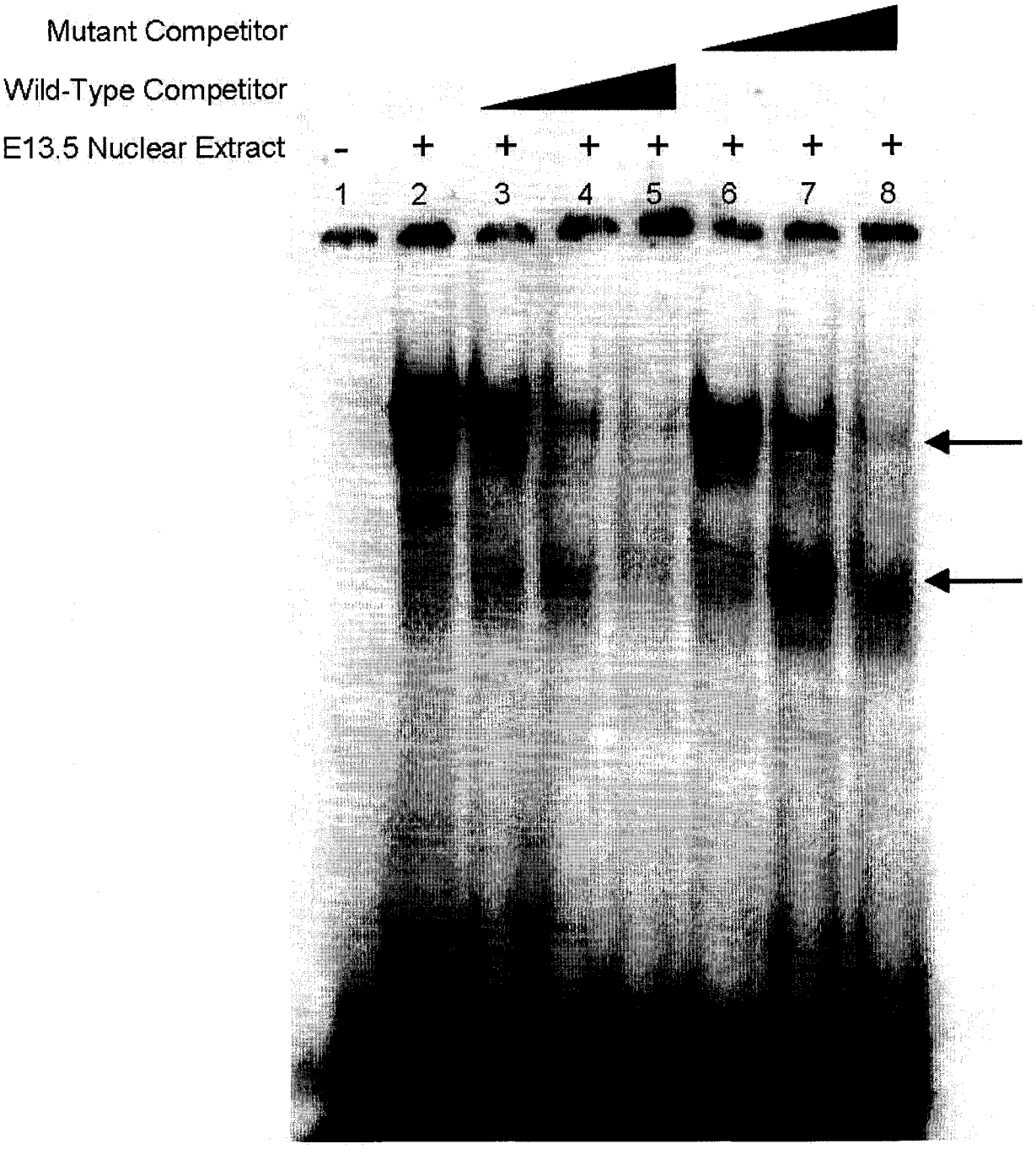
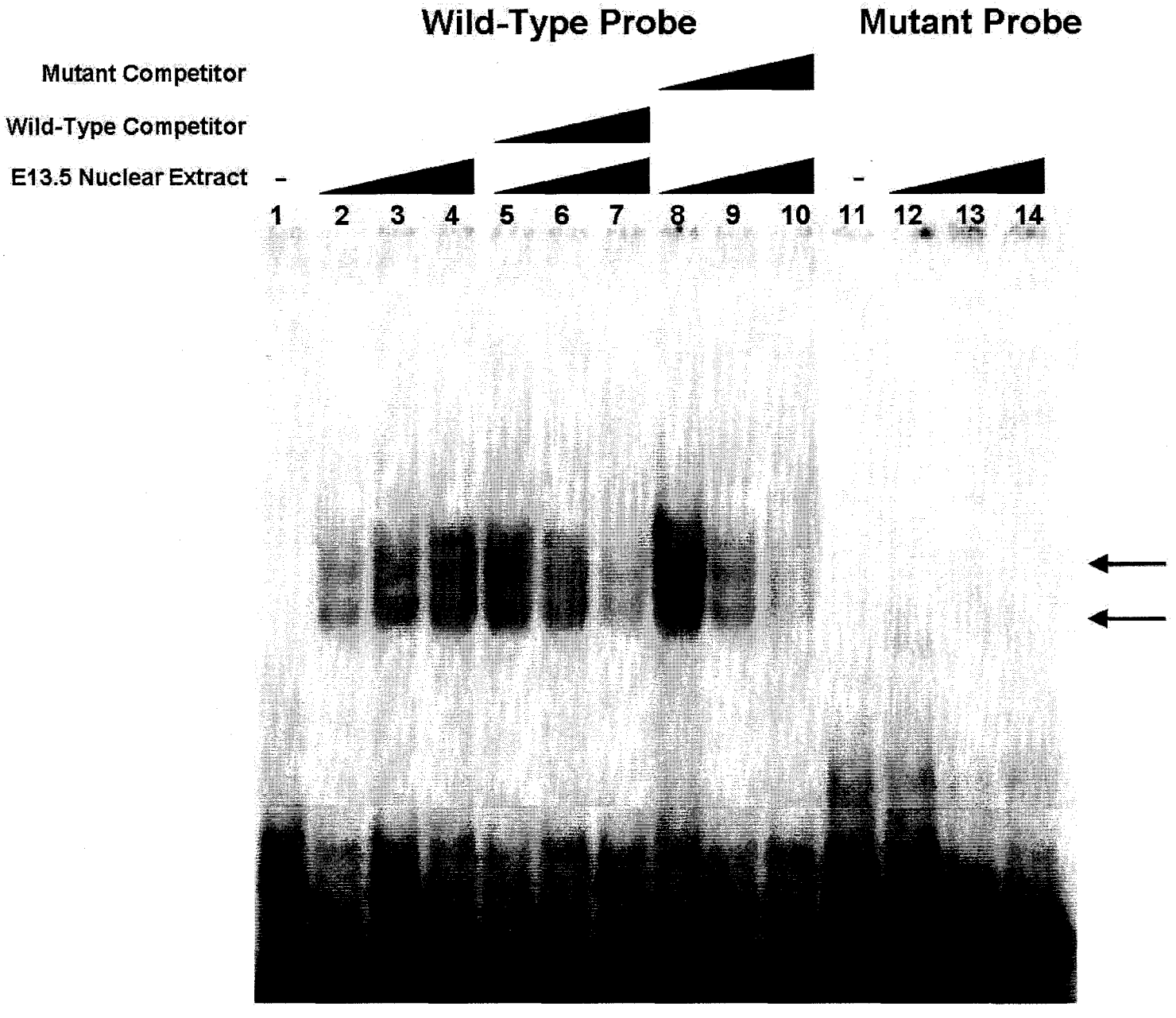

Mutant Competitor
Wild-Type Competitor
E13.5 Nuclear Extract
- + + + + + + +
1 2 3 4 5 6 7 8

Figure 3.11. *In vitro* analysis on the effect of mutagenesis of the putative homeodomain transcription factor binding site. An electrophoretic mobility shift assay with a radioactively labelled oligonucleotide fragment of the I56i enhancer and E13.5 forebrain nuclear extracts. Probes corresponding to a portion of the FP4 protected region sequence in wild-type form (lanes 1-10) or with the base-pair mutations corresponding to Mut2 (lanes 11-14) were used. The strength of the lower motility complexes with the Mut4 probe (lanes 12-14) is lower than that seen using the wild-type probe (lanes 2-4). However, competition with unlabelled oligonucleotide does not show a greater ability of either the wild-type or Mut4 oligonucleotides to block the binding of nuclear extracts to the DNA. The reduction in the complex of lower motility is similar with the wild-type competitor (lanes 5-7) in comparison to the Mut2 competitor (lanes 8-10). Increasing concentrations of nuclear extracts and unlabelled competitors are shown by increasing slopes. Lanes 5 and 8: 1:1 probe:competitor ratio; lanes 6 and 9: 10x molar excess competitor; lanes 7 and 10: 100x molar excess competitor. Data shown is representative of two repetitions.
Oligonucleotide sequences used: 4.4WT.For: 5'-TGTAGCCCGCTGATTACAGCGTTTTT-3'; 4.4WT.Rev: 5'-AAAAACGCTGTAATCAGCGGGCTACA-3'; 4.4Mut.For: 5'-TGTAG-CCCGCTCTAGACAGCGTTTTT-3'; 4.4Mut.Rev: 5'-AAAAACGCTGTCTAGAGCGGG-CTACA-3'


Wild-Type Probe
Mutant Probe
Mutant Competitor
Wild-Type Competitor
E13.5 Nuclear Extract
-
-
1 2 3 4 5 6 7 8 9 10 11 12 13 14

4. DISCUSSION AND CONCLUSIONS

4.1 Analysis of the Effects of Mutagenesis on I56i Activity in the Forebrain

4.1.1 Mutagenesis of the Protected Region FP1 (Mut1)

4.1.1a Activity of the Mut1 I56i Enhancer in Transgenic Mice

Mutagenesis of the I56i enhancer at the site of FP1, the putative GATA-1 binding site, appears to have a minimal effect on the activity of the enhancer in the transgenic lines produced. The activity of the enhancer in both lines is similar to the activity seen in the wild-type transgenic embryos. However, the primary transgenic embryos show a contradictory result with weaker activity in both the telencephalon and diencephalon of the mutants. Therefore, no conclusions can be made as to the effect of the mutation of this binding site within the I56i enhancer. The production of more transgenic embryos could potentially allow for the development of a more consistent pattern of activity and the effect of mutagenesis of the FP1 binding site on enhancer activity in the forebrain could hopefully be more accurately determined.

4.1.1b Evaluation of the Effect of the Mut1 Mutation on EMSA Experiments

The results of the electrophoretic mobility shift assays do not show a definitive effect following mutagenesis of the putative FP1 binding site. Although the complex of lower motility is weaker using the mutant probe, there is no difference seen between the wild-type and mutant competitors. This variability prevents any conclusions from being drawn regarding the effect of the base-pair mutations in this putative binding site. Thus it cannot be determined if the ability of the DNA-binding proteins found in the forebrain to bind the I56i enhancer is affected by the mutations introduced in the FP1 site.

4.1.1c Expression Patterns of Potential Transcription Factors Binding Within FP1

A putative GATA-1 binding site was selected as a candidate for mutagenesis within the protected area of FP1. The proteins of the GATA family encode transcription factors that are expressed in hematopoietic cell lineages. GATA factor expression is also seen in the developing heart and other organs (Morceau, Schnekenburger et al. 2004). Other than expression in hematopoietic progenitors, GATA factor expression does not overlap with significant areas of *Dlx* expression. However, since no transcription factor binding sites with a closer expression pattern to *Dlx* were originally identified in the protected region of FP1, the putative GATA-1 binding site was selected for further study.

A putative binding site for a MEIS1B/HOXa9 heterodimer was also identified in the FP1 region of the I56i enhancer. This binding site is located about 15 base pairs from the GATA-1 binding site discussed above. In mice, *Meis1* has been shown to be expressed within the telencephalon at embryonic stage E10.5, concurrent with *Dlx* expression (Toresson, Parmar et al. 2000). Since *Gata-1* shares minimal areas of overlapping expression domains with *Dlx*, it is not surprising that mutagenesis for the binding site of the GATA-1 protein does not reveal a significant difference in I56i activity. However, since a MEIS1b/HOXa9 binding site within the FP1 protected region has since been identified, further study into the role of the FP1 region in I56i activity may be warranted, given the overlapping areas of expression between *Meis1b/Hoxa9* and *Dlx*. Based on this, and the link between *Hox* and *Dlx* genes, mutagenesis of this putative MEIS1b/HOXa9 binding site may reveal a greater effect on the activity of the I56i enhancer in the forebrain.

4.1.2 Mutagenesis of the Protected Region FP2 (Mut2)

4.1.2a Activity of the Mut2 I56i Enhancer in Transgenic Mice

Overall, the greatest variation in the activity of the I56i enhancer in the forebrain of the transgenic mice was seen with the Mut2 construct. In one of the two transgenic lines and the one primary embryo obtained, the activity of the construct was greatly reduced in the telencephalon and nearly completely abolished in the diencephalon. This suggests that the base pair mutations introduced into the enhancer significantly reduce the ability of proteins to bind to the enhancer in these areas. This subsequently prevents the activation of the reporter construct and would similarly result in a decrease in enhancer activity in the forebrain. If protein binding in the region of FP2 is prevented, this may affect the ability of the enhancer to regulate its target genes, leading to potential developmental effects downstream. Although two of the three transgenic embryos show significant reductions in the two domains within the forebrain, the second transgenic line shows no effect at the whole mount level and a very restricted effect within the telencephalon when sectioned. This variability again limits the conclusions that can be drawn on the effect of the putative binding site on enhancer activity.

4.1.2b Evaluation of the Effect of the Mut2 Mutation on EMSA Experiments

Of the mutagenized protected regions investigated, the putative Engrailed binding site (Mut2) shows the most consistent reduction in the ability of proteins to bind the I56i enhancer in the EMSA experiments. Mutagenesis of the putative binding site appears to prevent protein binding on the DNA of the I56i enhancer. Since the base-pair mutations introduced may impact both a putative Engrailed or HMX/NKX5-1 binding site, it is possible that one or both of these proteins may be involved in the activation of the I56i enhancer. Further experiments would be required to confirm the identity of the proteins affected by

mutagenesis. In combination with the results seen in the transgenic experiments for Mut2, and if the production of more transgenic mice should result in a similar pattern of activity already seen, this suggests that transcription factors binding to a putative binding site within the FP2 region play an important role in the regulation of I56i activity.

4.1.2c Expression Patterns of Potential Transcription Factors Binding Within FP2

A putative binding site for Engrailed-1 was identified within the protected region FP2. Although the *Engrailed* genes do not show areas of expression consistent with the *Dlx* genes, they are active in the developing brain. The *Engrailed* genes have been shown to play a role in the patterning of the midbrain/hindbrain boundary. They are also important in the development and differentiation of the mesencephalic dopaminergic neurons (Simon, Thuret et al. 2004). At the time the areas for mutagenesis were selected, no other factors active in the forebrain were identified as potentially binding within the FP2 region. Therefore, given the role of Engrailed in the developing central nervous system it was selected as the focus for mutagenesis within the FP2 protected region.

Following mutagenesis experiments, a binding site for another homeodomain transcription factor, HMX3/NKX5-1, was identified in the protected region FP2. Expression of *Hmx3/Nkx5-1* has been seen in the developing inner ear, including precursors to the otic vesicle, a region where *Dlx* expression is also seen (Wang and Lufkin 2005). *Hmx* genes are also expressed in the first and second branchial arches as well as areas of the developing nervous system (Wang, Lo et al. 2000). This overlap in expression suggests that HMX3/NKX5-1 may be a more likely candidate for interacting with the I56i enhancer in regulating *Dlx* expression. However, the putative binding sites for both Engrailed and HMX3/NKX5-1 share a common core site and thus the base-pair mutations introduced would

have an impact on the potential binding of both proteins. Further experiments, such as repeating EMSA with a pure protein extract or an antibody to one of the candidate proteins, would be required to determine if the effects of mutagenesis are related to either or both of these proteins

4.1.3 Mutagenesis of the Protected Region FP3 (Mut3)

4.1.3a Activity of the Mut3 I56i Enhancer in Transgenic Mice

The three primary transgenic embryos obtained for the Mut3 construct each show a different level of activity in the forebrain. This lack of consistency again makes it difficult to draw any conclusions as to the effect of the mutagenesis of the putative DLX binding site on the activity of the I56i enhancer. However, previous research has already shown that DLX does bind to two areas within the I56i enhancer (Zerucha, Stuhmer et al. 2000; Zhou, Le et al. 2004). Therefore, mutagenesis of this binding site should have an effect on enhancer activity. However, the second putative DLX binding site within FP5 protected region may mask the effect of mutagenesis of the putative DLX binding site within the FP3 region.

The similarity between the DLX binding sites within the I56i enhancer and the I12b enhancer in the *Dlx1/Dlx2* locus can also be used to predict the nature of the role of DLX in the I56i enhancer (see also section 4.8). Mutagenesis of the putative DLX binding sites in I12b resulted in a large reduction of enhancer activity in the telencephalon and a reduction in the diencephalon (Poitras, Ghanem et al. 2007). This supports the loss of activity seen in two of the three transgenic embryos obtained for mutagenesis of the DLX binding sites within the I56i enhancer.

4.1.3b Evaluation of the Effect of the Mut3 Mutation on EMSA Experiments

Mutagenesis of the putative DLX binding site within the protected region FP3 appeared to affect the ability of proteins to bind the I56i enhancer. Although the results are not consistent throughout all experiments, there is a general decrease in the intensity of the band of lower motility using the mutant probe. Since a putative DLX binding site was mutated in this region, the decrease in the intensity of the band suggests that mutagenesis is preventing DLX from binding to the enhancer DNA in this region. Although other transcription factors may also be binding to this putative binding site, previous experiments have demonstrated that DLX binds to the I56i enhancer within the protected region FP3 and that mutagenesis of this site impairs the ability of DLX to bind this fragment of the I56i enhancer (Zerucha, Stuhmer et al. 2000; Zhou, Le et al. 2004). This provides support to the theory that it is the loss of DLX binding in this region that is causing a decrease in the intensity of DNA-protein complex.

4.1.3c Expression Patterns of Potential Transcription Factors Binding Within FP3 and FP5

Although DLX is the primary candidate for binding within the FP3 (and FP5) putative binding sites, a number of other homeodomain proteins were also identified within the FP3 region. Of these transcription factors identified, only the GSH proteins share areas of expression in the forebrain similar to that of the DLX proteins and the I56i enhancer. Both GSH-1 and GSH-2 have been show in the developing mouse forebrain, including the developing LGE and MGE (Szucsik, Witte et al. 1997; Li, Schrick et al. 1999). Given this overlapping expression domain, the GSH proteins present another set of potential candidates for binding the I56i enhancer within the FP3 protected region and influencing the activity of the enhancer. The GSH proteins share a similar core binding site to the DLX core selected

for mutagenesis and thus would be similarly affected by the mutagenesis experiments. Therefore, the effects seen in the EMSA experiments and transgenic animals may also be attributed to the GSH proteins. If further research confirms that mutagenesis of this site causes a significant impact on the activity of the I56i enhancer, experiments would be required to determine if this was caused by an impact on DLX or GSH binding, or both.

For both the FP3 and FP5 protected regions, the other transcription factors potentially binding within these regions do not demonstrate expression patterns that overlap as extensively with DLX expression or I56i enhancer activity. A binding site for HMX3 is found within the FP3 protected region. Murine *Hmx3* expressions is found in areas that overlap with *Dlx* gene expression, such as the first and second branchial arches and otic vesicle, but is not found in the forebrain (Wang, Lo et al. 2000). Other proteins with putative binding sites within the FP3 region include NKX2.5/CSX (cardiac precursors), CART-1 (found in chondrocytes), and S8 (in mesodermal precursors) (de Jong, van der Heijden et al. 1993; Zhao, Zhou et al. 1993; Wang, Lo et al. 2000; Akazawa and Komuro 2005). Putative binding sites for PDX-1 (found in pancreatic precursors), GATA-2 (hematopoietic progenitors) LIM3 (expressed in the spinal cord, lungs, and pituitary gland), and S8 were identified in the FP5 protected region (de Jong, van der Heijden et al. 1993; Morceau, Schnekenburger et al. 2004; Wilding and Gannon 2004; Harigae 2006; Mullen, Colvin et al. 2007). Although it is more likely that DLX or GSH are interacting with the I56i enhancer in the protected regions of FP3 and FP5, the role of those other transcription factors must be investigated if both DLX and GSH are not the main transcription factors binding to the FP3 and FP5 regions.

4.1.4 Mutagenesis of the Protected Region FP4 (Mut4)

4.1.4a Activity of the Mut4 I56i Enhancer in Transgenic Mice

The majority of the primary transgenic embryos obtained carrying the Mut4 construct show only very faint reporter activity in the forebrain. It is possible that this residual activity of the transgene is not related to the wild-type activity of the I56i enhancer. However, each of the embryos identified as a primary transgenic embryo had a staining pattern consistent with at least one area of endogenous I56i activity, such as the AER. Further, the PCR results from the genomic DNA were able to provide an estimate of the number of primary embryos expected for each staining procedure. As a result, it is reasonable to assume that the faint staining is related to the presence of the construct in these embryos.

The drastic effects on enhancer activity seen in all of the primary embryos produced suggest that the putative homeodomain site is significant for the functional activity of the I56i enhancer. The reduced binding ability seen after introduction of the base-pair mutations suggests that a protein binding in this region is required for the activation or continued function of the enhancer. To get a better understanding of the role this factor plays in the regulation of the enhancer, the identification of this factor is required.

4.1.4b Evaluation of the Effect of the Mut4 Mutation on EMSA Experiments

The EMSA experiments conducted using probes containing the base pair mutations introduced into the protected region FP4 did not reveal a consistent effect on the formation of DNA-protein complexes. Proteins from the nuclear extract did not appear to bind as well to the mutant binding site when compared to the wild-type binding site. However, the use of the unlabelled mutant oligonucleotide as a competitor showed the same disruption in DNA-protein binding as was seen when using the wild-type oligonucleotide competitor. Therefore,

the results of the EMSA experiments were not consistent enough to draw any conclusions. This may be due to the fact that the binding site selected for mutagenesis was a general TAAT homeodomain core. Proteins with a low binding specificity for this site may still be able to bind even after mutagenesis of the putative binding site.

4.1.4c Expression Patterns of Potential Transcription Factors Binding Within FP4

At the time a putative binding site was selected for mutagenesis in the protected region of FP4, no known transcription factor binding sites had been identified in this region. Therefore, the general homeodomain site, TAAT, was selected. Since this is the core of many homeodomains, mutagenesis of this site could influence a wide range of potential targets. However, since the completion of mutagenesis, a putative binding site for HMX2/NKX5-2 was identified just five base pairs from the TAAT selected for mutagenesis. As with *Hmx3* described above, *Hmx2* is expressed in the first and second branchial arches and the otic vesicle, areas of *Dlx* expression, as well as the developing hypothalamus (Wang, Lo et al. 2000; Wang and Lufkin 2005). Since this protein is found in areas where the I56i enhancer is active, it may be another potential candidate for influencing the activity in the I56i enhancer and could be investigated further in the future.

4.1.5 Mutagenesis of the Protected Regions FP5-FP8 (Mut5 – Mut8)

4.1.5a Prioritization of Mutant Construct for Production of Transgenic Mice

In total, up to eight putative binding sites were identified within the I56i enhancer. However, only four (FP1-FP4) have been studied to date through mutagenesis and electrophoretic mobility shift assays. The focus for the production of transgenic animals was limited to three of the putative binding sites identified by footprint analysis. FP1, FP2 and FP4 were selected as they were clearly identified on the original footprint experiments.

Footprint site 3, the putative DLX site, was originally used only to test the production of the mutants since this site, along with the second putative DLX binding site (FP5) had already been studied (Zerucha, Stuhmer et al. 2000).

After the identification of footprint sites 6, 7, and 8 and the selection of putative binding sites for mutagenesis, evaluation through transgenic mice and EMSA could have been performed for these footprint sites. However, given the ambiguous results already obtained for the initial protected regions investigated it was decided to focus on the original sites. Further, the majority of the original sites were already linked to a potential binding protein and were thus more likely to yield significant results.

4.1.5b Expression Patterns of Potential Transcription Factors Binding Within FP6-FP8

Originally, no transcription factor binding sites for proteins expressed in regions of I56i activity were identified in the protected regions of FP6-FP8. Areas for mutagenesis were selected based on TAAT homeodomain cores and conserved regions. However, since the selection of these sites for mutagenesis, transcription factor binding sites for genes expressed in regions of I56i activity have been identified within the protected regions of FP6 and FP7.

A binding site for TST-1/OCT-6 was identified in the region for FP6 and the core of this binding site is identical to the TAAT selected previously for mutagenesis in this region. Expression of *Tst-1* is consistent with some areas of endogenous *Dlx* expression. It is seen in areas of the forebrain and areas of the olfactory bulb (Alvarez-Bolado, Rosenfeld et al. 1995).

Just outside of the base pair mutations introduced into the protected region of FP7, a binding site for BRN-2 was identified. This protein is found in the forebrain, including both the telencephalon and diencephalon, during mouse development. This is consistent with the

pattern of expression seen for *Dlx* (Alvarez-Bolado, Rosenfeld et al. 1995). If the putative binding site already mutagenized within protected region FP7 does not show any effect on I56i activity, changing the focus to the area of the BRN-2 binding site may show more of an effect given the similarity of the expression domains of *Brn-2* and *Dlx*.

4.2 Summary of Mutagenesis Results on I56i Activity in the Forebrain

MatInspector software revealed the presence of a putative DLX binding site within FP5 as well as the potential binding of other factors within this region (see section 4.1.3c). Overall, the results of mutagenesis of putative binding sites within the I56i enhancer sequence on I56i activity in the forebrain were inconclusive. The activity of the mutated enhancer varied among the transgenic embryos produced for each construct. This was seen amongst all four constructs in which transgenic mice were studied. Although some transgenic lines or primary embryos showed a large reduction in forebrain activity, other transgenic lines or primary embryos showed no changes. This is especially evident for the Mut3 construct, in which each of the primary transgenic embryos shows a different degree of I56i activity, ranging from faint activity to activity similar to that seen with the wild-type enhancer. Because of this range in activity within each of the constructs, no conclusions can be drawn from the transgenic mice produced in terms of the effect of mutagenesis of the putative binding sites on enhancer activity in the forebrain.

4.3 Analysis of the Effect of Mutagenesis on I56i Activity in the Branchial Arches

In total, for all of the transgenic experiments, four of the five transgenic lines produced and all primary transgenic embryos showed a significant reduction or a complete loss of construct activity in the branchial arches. This suggests that there are multiple binding sites within the I56i enhancer that are critical for the proper activity of the enhancer in the

branchial arches. Mutagenesis of these binding sites prevents proteins from binding and limits the activity of the enhancer. However, within each of the constructs, there is some variability in the degree to which branchial arch activity is affected. For example, within the Mut1 construct, both primary embryos and one of the two transgenic lines show a reduction of activity in the hyoid arch and a loss or reduction in the mandibular arch. However, the other transgenic lines show activity in these regions similar to that seen in the wild-type. As a result, it is difficult to make definitive conclusions on the effect of the mutagenesis.

One reason for the possible variability in the activity seen in the branchial arches could be due to the ability of the reporter construct to be activated within this area. In the original I56i transgenic mice produced (13 primary embryos and 4 transgenic lines) only half (9/17 = 53%) of these transgenic animals showed activity in the branchial arches (Zerucha, Stuhmer et al. 2000). However, within the transgenic embryos that did show activity in the branchial arches, the activity was consistent. Since only half of the transgenic lines produced with the I56i construct show activity in the branchial arches, a greater number of transgenic lines are needed to accurately evaluate the activity of the enhancer in this region.

4.4 Variability within Transgenic Animals Limits Evaluation of the Effect of Mutagenesis

The variability seen in the transgenic animals produced to date prevent a complete evaluation of the effect of the introduced mutations on activity of the I56i enhancer. In order to get a more consistent and precise representation of the true activity of the enhancer containing the mutated putative binding sites, more transgenic mice would need to be produced. At the time of the original injection, significant difficulty in the injection process was resulting in poor efficiency and low integration rates. As a result, and in association with

the cost of injection and raising the mice, only a limited number of injections were completed for the constructs discussed above. If more transgenic animals could be produced for each of the constructs, it could be expected that a more accurate representation of the true activity of each of the mutant constructs could be obtained.

4.5 Potential Causes of Transgenic Variability

There are a number of potential causes for the variability seen in the transgenic mice produced. First, a variation in the number of copies of the transgene construct that integrated into the genome following injection into fertilized eggs may have caused inconsistencies in the results obtained. This may occur if multiple copies of the transgene integrate into different areas of the genome or if a concatemer of multiple repeats of the transgene integrates into the genome. If multiple copies of the construct are integrated into the genome, the amount of the β-galactosidase enzyme produced will be greater than in a single integration event. As a result, the amount of blue staining product produced will also be increased and a difference in the perceived activity of the construct will result. In order to account for this potential problem, an analysis of gene copy number can be completed using real-time PCR. This would allow the number of integrated transgenes to be determined and allow for a more accurate comparison between the transgenic lines

Another potential explanation for the variability seen in β-galactosidase staining in the transgenic lines is the position effect, which is a result of the location of the integration of the reporter construct (Alberts 2002). If the construct is integrated in an area of the genome with neighbouring enhancers, these enhancers may interact with the reporter gene causing a variation in the production of β-galactosidase than what would normally be expected. An interaction with surrounding heterochromatin may also cause this position effect. This often

results in a reduction in the number of cells expressing the reporter gene due to the compact nature of the chromosomes and the subsequent reduction in gene expression seen due to the inability to initiate transcription. With the mutant constructs, this effect would increase the perceived effect of the mutation of the putative binding site on enhancer activity. A potential solution to this is to include insulator elements in the constructs. Specialized proteins binding to the insulator DNA sequences prevent the position effect by blocking the interaction of neighbouring enhancers and heterochromatin (Alberts 2002). Including insulator elements in the reporter constructs would eliminate the artificial decrease in activity seen due to the presence of heterochromatin and eliminate the effect of any neighbouring enhancer elements located in the region of integration. This could potentially lead to more reliable and consistent results in the transgenic animals produced.

Finally, an integration effect may result in activity seen in areas of non-endogenous I56i activity. If a copy of the construct integrates into a transcriptionally active region of DNA, the construct will be transcribed without being activated by the β-globin promoter. This will not be representative of the activity of the construct itself but of the transcriptionally active region into which the construct was introduced. This is likely the cause of the excess staining seen in a number of primary embryos and line7838 for the Mut1 construct. In this line, staining can be seen in regions of the brain and skull not seen in wild-type I56i activity. A copy of the construct has likely integrated into a region of DNA that is transcribed in these structures. This is not representative of endogenous I56i activity or *Dlx* expression. Insulators can again be used to prevent the integration effect as they would prevent the construct from being influenced by the regulatory activity of surrounding genes.

This would again provide more reliable results by eliminating the ectopic expression seen from the active transcription of the neighbouring gene.

4.6 EMSA Supports Presence of Transcription Factor Binding Sites within the I56i Enhancer

Consistent with the results of previous studies that show proteins expressed in the brain and branchial arches bind the I56i enhancer, the EMSA experiments show that proteins in the nuclear extracts from the brain (E11.5), forebrain (E13.5) and branchial arches (E13.5) are all able to bind to the enhancer (Zerucha, Stuhmer et al. 2000; Zhou, Le et al. 2004). This confirms that proteins present within these tissues have the capacity to bind to the enhancer and potentially play a role in regulating the activity of the I56i enhancer. In the case of EMSA experiments completed with the oligonucleotide probes, it also confirms that proteins bind within those specific 26 base pair regions. Although this binding cannot be specifically localized to the nucleotides targeted for mutation, it does suggest that there is a site within these small regions in which transcription factors are able to bind. This supports the belief that binding sites exist within the areas of the enhancer identified by *DNase*I footprinting. This is consistent with the results in previous studies which also show that proteins are able to bind the I56i enhancer (Zerucha, Stuhmer et al. 2000).

Overall, the mutagenesis of a putative binding site within the FP2 protected region shows the most consistent effect in the EMSA experiments. Mutagenesis of this site tends to reduce the ability of proteins to bind the DNA. This suggests that the mutagenesis of the binding site identified within the FP2 region prevents proteins from binding onto a fragment of DNA on which they would normally be able to bind. MatInspector software has already provided two potential proteins that may be binding this site, Engrailed-1 and HMX3/NKX5-

1. Further studies, such as more EMSA experiments using a pure protein extract from one of these candidate proteins or supershifts with an antibody to one of the proteins, would be required to more precisely identify which protein is binding in the region of FP2. Even though some differences are seen *in vitro*, this variation may not be sufficient to cause a significant variation *in vivo*. The results from EMSA experiments must be combined with results from transgenic experiments to determine the role of the putative binding sites in activity of the I56i enhancer.

4.7 Potential Causes of Low Resolution in EMSA Experiments

The most likely cause of the poor resolution within the EMSA experiments is poor quality of the nuclear protein extracts. The nuclear protein extracts were prepared from the dissected forebrains and brains of E13.5 and E11.5 mouse embryos. The preparation from these structures may have included some tissue from surrounding regions. Furthermore, to increase the amount of tissue collected, nuclear protein extracts were prepared from the head of E11.5 embryos as opposed to the dissected forebrain. This would result in a greater number of proteins being present in the extract than would normally be found in the tissues of interest. A greater number of total proteins in the extract may result in an increase in the number of proteins that may bind to the DNA probe, including DNA-binding proteins that would not normally be present in tissues where the I56i enhancer would be active. This could lead to a greater number of bands of lower motility being present on the gel. A lower resolution on the gel with the potential for smears could also occur due to the large number of proteins potentially binding to the DNA.

If a candidate protein could be identified as potentially binding to any of the putative binding sites identified within the I56i enhancer, it could be possible to use a pure protein

extract. This would ensure that only one protein was present during the experiment and thus only a single band of lower motility would be expected (assuming there is only one binding site within the DNA). Furthermore, if an antibody to the protein could be obtained, an EMSA experiment could be completed using the antibody to supershift the protein-DNA complex in order to confirm that the protein used is actually binding the DNA.

Another factor that may have resulted in poor results in the EMSA experiments is the degradation of the proteins in the nuclear extracts. Any errors in the storage of the tissues or the preparation of the nuclear extracts may have caused some protein degradation. This could reduce the affinity and specificity of the proteins for a specific binding site. Fewer bands would result if the affinity of the proteins is reduced and do not bind the DNA or a much larger number of bands, such as smearing, could result if the specificity of the protein is reduced and binds DNA at multiple non-specific locations.

Poor resolution of the gel may also be caused by the size of the DNA fragments used in the experiments. Fragments of 150-400 base pairs were used for a number of EMSA experiments. In combination with the binding of one protein, or multiple proteins, the size of the DNA-protein complex can become quite large. This would reduce the ability of the complex to travel through the gel. This effect could be quite variable depending on the size of the complexes and the permeability of the gel matrix. The larger fragments used also require a longer migration time, which could also result in further protein degradation and the breakdown of the DNA-protein binding complex. These could all be causes of the smearing seen in the EMSA experiments completed using large fragments or the entire I56i sequence.

4.8 Extrapolation from a Closely Related Regulatory Element, I12b

Although the results of this study are not consistent enough to draw any final conclusions, some similarities can be drawn between the results discussed here and those seen following the mutagenesis of the I12b enhancer. Poitras et al. identified putative binding sites within the I12b enhancer element, mutagenized these sites to determine their effect on the activity of the I12b enhancer, and identified transcription factors binding to this enhancer (Poitras, Ghanem et al. 2007). This research was able to identify transcription factors, such as MASH1 and MEIS2, that play a role in the activity of the I12b enhancer and its regulation of *Dlx* expression. Given that the *Dlx* bigene loci are believed to have evolved from the duplication of an original *Dlx* cluster, it can be suggested that the role of the *Dlx5/Dlx6* intergenic enhancer I56i may be similar to that of the *Dlx1/Dlx2* intergenic enhancer I12b. Since some of the results seen for the I56i enhancer element recapitulate what was seen following mutagenesis of the I12b enhancer, further evaluation of the factors affecting I56i activity is warranted.

4.9 Potential Role of the I56i Enhancer Element in Regulating *Dlx* Expression

Results of the transgenic and EMSA experiments suggest that alteration of the I56i enhancer sequence affect the affinity of DNA binding proteins for the enhancer and the subsequent activation of the enhancer. However, given the variability of the results, the precise impact of the introduced mutations cannot be determined.

Previous research has shown that DLX binds the I56i enhancer (Zerucha, Stuhmer et al. 2000). The experiments in the current study have also shown that other factors, such as Engrailed-1 or HMX3/NKX5-1, are also likely able to bind the enhancer and play a role in the regulation of its activity. It has been suggested that, due to its location in the intergenic

region between *Dlx5* and *Dlx6*, the I56i enhancer may have a function in the regulation of the closely linked *Dlx5* and *Dlx6* genes. Although this cannot be confirmed via this research, the factors that bind to the enhancer may play a role in the regulation of the *Dlx* genes. DLX has been shown to be able to form dimers with at least one other transcription factor, MSX (Zhang, Hu et al. 1997). Since it has already been shown that DLX binds the I56i enhancer, it is possible that the other transcription factors identified may bind to the enhancer via this interaction. Poitras et al. (2007) suggested that activation of the I12b enhancer occurs through a number of potential factors but DLX is required to maintain activity (Poitras, Ghanem et al. 2007). It is possible that a similar regulatory scheme is operating in I56i. One or more of the transcription factors identified may be required for the activation of the enhancer but DLX will maintain this activity past a certain stage. This complex regulatory mechanism may account for the limited effects seen in the mutagenesis experiments. Other transcription factors may be compensating for the reduction in binding of the transcription factor affected by the mutagenesis.

4.10 Possible Connection Between the I56i Enhancer and Autism

The *Dlx* genes have been shown to play a significant role in development. The I56i enhancer, believed to be involved in the regulation of the *Dlx* genes, has also been linked to a serious human condition. Autism, or autistic spectrum disorder (ASD) is a potentially severe syndrome that affects communication and social skills in humans. Family and twin studies have shown that there is likely a genetic link to autism. It has also been suggested that some forms of autism may be a result of ineffective inhibition of excitatory neurons in the brain. Since *Dlx1*, *Dlx2*, *Dlx5* and *Dlx6* play a role in the development of inhibitory interneurons in the forebrain, variation in these genes have been studied as a potential cause of autism. One

area of interest is on chromosome 7q, the same chromosome that contains *Dlx5* and *Dlx6* (Hamilton, Woo et al. 2005).

In a comparison study of autism probands and non-autistic siblings, a total of thirty-three variants within the *Dlx1*, *Dlx2*, *Dlx5* and *Dlx6* coding and non-coding sequences were observed. One of these variants was a non-synonymous single nucleotide polymorphism (SNP) located within the I56i enhancer. This SNP results in the change of an adenine nucleotide for a guanine nucleotide close to one of the two putative *Dlx* binding sites found in the I56i sequence. Given that the I56i sequence is ultraconserved, the presence of a SNP suggests that this variation may play a potential role in the development of autism.

The role of this SNP has been studied by another member of our laboratory (Luc Poitras, unpublished results). DNA binding assays showed that the ability of at least one forebrain transcription factor to bind the I56i sequence was impaired by the presence of the SNP. Furthermore, activity of the I56i enhancer containing the SNP in a *lacZ* reporter construct was reduced in some areas of the telencephalon. Further work is being conducted to determine the effect of this variation on *Dlx* expression and the development of the forebrain.

Even though the Engrailed proteins have not been identified in regions of I56i activity, both the *Engrailed* genes and I56i have been linked to the study of autism. As discussed above, a relationship between a SNP in the I56i enhancer and autism has been suggested. Interestingly, a link between autism and the *Engrailed* genes has also been suggested (Murcia, Gulden et al. 2005). The putative Engrailed-1 binding site is only six base pairs upstream of the SNP in I56i. Given the close proximity of these two areas of interest, further study into the role of the I56i enhancer in the development of autism may be warranted.

4.11 Future Directions

Given that the results of the transgenic mice experiments were so variable, a potential first step would be to repeat the injection of each of the constructs to increase the number of transgenic mice and, ideally, obtain more consistent results. With a greater number of transgenic mice for each construct, a more consistent pattern of activity should emerge in the majority of the embryos. With this, the minority of embryos with a different pattern of activity can be considered outliers with more confidence. The new constructs created for the protected regions of FP6, FP7, and FP8 can also be injected to determine if these mutations have any effect on the activity of the I56i enhancer. Since putative binding sites for some transcription factors sharing expression patterns with *Dlx* have been identified in these regions, mutagenesis may reveal that these binding sites have a significant function in *Dlx* regulation. Other sites, such as FP8, may not show any effect since no known transcription factors have been identified as binding within the protected region.

To increase the accuracy of the result seen in the transgenic mice, insulators can be added to each of the constructs. This would prevent some of the potential variation seen in the transgenic animals and aid in obtaining more consistent results. Finally, for the transgenic lines already produced, an evaluation into the number of copies of the construct inserted can be performed by real-time PCR. This would aid in the evaluation of the transgenic animals already produced. If the transgenic lines showing strong enhancer activity have a greater number of copies of the constructs inserted, it can be inferred that the strong activity is related to the larger number of copies. Thus, it can be predicted that the true activity of the construct would be significantly less and similar to the other transgenic mice with weaker activity.

Finally, if the mutagenesis of a putative binding site shows a significant effect in both the transgenic mice and EMSA experiments, it will be important to identify the protein or proteins binding within this region. If a candidate protein is determined, supershift EMSA experiments using an antibody to this protein can be performed. If possible, EMSA experiments using a pure protein population can be completed. Each of these experiments would aid in identifying the protein binding to the labelled probe. If no candidate protein is known, the band of lower motility may be isolated from an EMSA experiment and mass spectrometry may be used to identify the isolated protein.

5. REFERENCES

Acampora, D., G. R. Merlo, et al. (1999). "Craniofacial, vestibular and bone defects in mice lacking the Distal-less-related gene Dlx5." Development **126**(17): 3795-809.

Ahituv, N., Y. Zhu, et al. (2007). "Deletion of ultraconserved elements yields viable mice." PLoS Biol **5**(9): e234.

Akazawa, H. and I. Komuro (2005). "Cardiac transcription factor Csx/Nkx2-5: Its role in cardiac development and diseases." Pharmacol Ther **107**(2): 252-68.

Alberts, B. (2002). Molecular biology of the cell. New York, Garland Science.

Alvarez-Bolado, G., M. G. Rosenfeld, et al. (1995). "Model of forebrain regionalization based on spatiotemporal patterns of POU-III homeobox gene expression, birthdates, and morphological features." J Comp Neurol **355**(2): 237-95.

Amores, A., A. Force, et al. (1998). "Zebrafish hox clusters and vertebrate genome evolution." Science **282**(5394): 1711-4.

Anderson, S. A., D. D. Eisenstat, et al. (1997). "Interneuron migration from basal forebrain to neocortex: dependence on Dlx genes." Science **278**(5337): 474-6.

Bejerano, G., M. Pheasant, et al. (2004). "Ultraconserved elements in the human genome." Science **304**(5675): 1321-5.

Bendall, A. J. and C. Abate-Shen (2000). "Roles for Msx and Dlx homeoproteins in vertebrate development." Gene **247**(1-2): 17-31.

Beverdam, A., G. R. Merlo, et al. (2002). "Jaw transformation with gain of symmetry after Dlx5/Dlx6 inactivation: mirror of the past?" Genesis **34**(4): 221-7.

Bulfone, A., H. J. Kim, et al. (1993). "The mouse Dlx-2 (Tes-1) gene is expressed in spatially restricted domains of the forebrain, face and limbs in midgestation mouse embryos." Mech Dev **40**(3): 129-40.

Cobos, I., V. Broccoli, et al. (2005). "The vertebrate ortholog of Aristaless is regulated by Dlx genes in the developing forebrain." J Comp Neurol **483**(3): 292-303.

Cohen, S. M. (1990). "Specification of limb development in the Drosophila embryo by positional cues from segmentation genes." Nature **343**(6254): 173-7.

Cohen, S. M., G. Bronner, et al. (1989). "Distal-less encodes a homoeodomain protein required for limb development in Drosophila." Nature **338**(6214): 432-4.

Cohen, S. M. and G. Jurgens (1989). "Proximal-distal pattern formation in Drosophila: cell autonomous requirement for Distal-less gene activity in limb development." Embo J **8**(7): 2045-2055.

de Jong, R., J. van der Heijden, et al. (1993). "DNA-binding specificity of the S8 homeodomain." Nucleic Acids Res **21**(20): 4711-20.

Depew, M. J., J. K. Liu, et al. (1999). "Dlx5 regulates regional development of the branchial arches and sensory capsules." Development **126**(17): 3831-46.

Depew, M. J., T. Lufkin, et al. (2002). "Specification of jaw subdivisions by Dlx genes." Science **298**(5592): 381-5.

Depew, M. J., C. A. Simpson, et al. (2005). "Reassessing the Dlx code: the genetic regulation of branchial arch skeletal pattern and development." J Anat **207**(5): 501-61.

Dolle, P., M. Price, et al. (1992). "Expression of the murine Dlx-1 homeobox gene during facial, ocular and limb development." Differentiation **49**(2): 93-9.

Eisenstat, D. D., J. K. Liu, et al. (1999). "DLX-1, DLX-2, and DLX-5 expression define distinct stages of basal forebrain differentiation." J Comp Neurol **414**(2): 217-37.

Ellies, D. L., R. M. Langille, et al. (1997). "Specific craniofacial cartilage dysmorphogenesis coincides with a loss of dlx gene expression in retinoic acid-treated zebrafish embryos." Mech Dev **61**(1-2): 23-36.

Ellies, D. L., D. W. Stock, et al. (1997). "Relationship between the genomic organization and the overlapping embryonic expression patterns of the zebrafish dlx genes." Genomics **45**(3): 580-90.

Feledy, J. A., M. I. Morasso, et al. (1999). "Transcriptional activation by the homeodomain protein distal-less 3." Nucleic Acids Res **27**(3): 764-70.

Feng, J., C. Bi, et al. (2006). "The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator." Genes Dev **20**(11): 1470-84.

Force, A., M. Lynch, et al. (1999). "Preservation of duplicate genes by complementary, degenerative mutations." Genetics **151**(4): 1531-45.

Ghanem, N., O. Jarinova, et al. (2003). "Regulatory roles of conserved intergenic domains in vertebrate Dlx bigene clusters." Genome Res **13**(4): 533-43.

Ghanem, N., M. Yu, et al. (2007). "Distinct cis-regulatory elements from the Dlx1/Dlx2 locus mark different progenitor cell populations in the ganglionic eminences and different subtypes of adult cortical interneurons." J Neurosci **27**(19): 5012-22.

Griffiths, A. J. F. (2000). An introduction to genetic analysis. New York, W.H. Freeman.

Hamilton, S. P., J. M. Woo, et al. (2005). "Analysis of four DLX homeobox genes in autistic probands." BMC Genet **6**: 52.

Harigae, H. (2006). "GATA transcription factors and hematological diseases." Tohoku J Exp Med **210**(1): 1-9.

Hogan, B., F. Costantini, et al. (1986). Manipulating the mouse embryo : a laboratory manual. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory.

Katzman, S., A. D. Kern, et al. (2007). "Human genome ultraconserved elements are ultraselected." Science **317**(5840): 915.

Kuwajima, T., I. Nishimura, et al. (2006). "Necdin promotes GABAergic neuron differentiation in cooperation with Dlx homeodomain proteins." J Neurosci **26**(20): 5383-92.

Le, T. N., G. Du, et al. (2007). "Dlx homeobox genes promote cortical interneuron migration from the Basal forebrain by direct repression of the semaphorin receptor neuropilin-2." J Biol Chem **282**(26): 19071-81.

Li, H., J. J. Schrick, et al. (1999). "Novel strategy yields candidate Gsh-1 homeobox gene targets using hypothalamus progenitor cell lines." Dev Biol **211**(1): 64-76.

Liu, J. K., I. Ghattas, et al. (1997). "Dlx genes encode DNA-binding proteins that are expressed in an overlapping and sequential pattern during basal ganglia differentiation." Dev Dyn **210**(4): 498-512.

Long, J. E., S. Garel, et al. (2003). "DLX5 regulates development of peripheral and central components of the olfactory system." J Neurosci **23**(2): 568-78.

Masuda, Y., A. Sasaki, et al. (2001). "Dlxin-1, a novel protein that binds Dlx5 and regulates its transcriptional function." J Biol Chem **276**(7): 5331-8.

Maxam, A. M. and W. Gilbert (1980). "Sequencing end-labeled DNA with base-specific chemical cleavages." Methods Enzymol **65**(1): 499-560.

Morasso, M. I., A. Grinberg, et al. (1999). "Placental failure in mice lacking the homeobox gene Dlx3." Proc Natl Acad Sci U S A **96**(1): 162-7.

Morceau, F., M. Schnekenburger, et al. (2004). "GATA-1: friends, brothers, and coworkers." Ann N Y Acad Sci **1030**: 537-54.

Mullen, R. D., S. C. Colvin, et al. (2007). "Roles of the LHX3 and LHX4 LIM-homeodomain factors in pituitary development." Mol Cell Endocrinol **265-266**: 190-5.

Murcia, C. L., F. Gulden, et al. (2005). "A question of balance: a proposal for new mouse models of autism." Int J Dev Neurosci **23**(2-3): 265-75.

Panganiban, G. (2000). "Distal-less function during Drosophila appendage and sense organ development." Dev Dyn **218**(4): 554-62.

Panganiban, G. and J. L. Rubenstein (2002). "Developmental functions of the Distal-less/Dlx homeobox genes." Development **129**(19): 4371-86.

Park, B. K., S. M. Sperber, et al. (2004). "Intergenic enhancers with distinct activities regulate Dlx gene expression in the mesenchyme of the branchial arches." Dev Biol **268**(2): 532-45.

Pennacchio, L. A., N. Ahituv, et al. (2006). "In vivo enhancer analysis of human conserved non-coding sequences." Nature **444**(7118): 499-502.

Poitras, L., N. Ghanem, et al. (2007). "The proneural determinant MASH1 regulates forebrain Dlx1/2 expression through the I12b intergenic enhancer." Development **134**(9): 1755-65.

Qiu, M., A. Bulfone, et al. (1997). "Role of the Dlx homeobox genes in proximodistal patterning of the branchial arches: mutations of Dlx-1, Dlx-2, and Dlx-1 and -2 alter

morphogenesis of proximal skeletal and soft tissue structures derived from the first and second arches." Dev Biol **185**(2): 165-84.

Qiu, M., A. Bulfone, et al. (1995). "Null mutation of Dlx-2 results in abnormal morphogenesis of proximal first and second branchial arch derivatives and abnormal differentiation in the forebrain." Genes Dev **9**(20): 2523-38.

Ryoo, H. M., H. M. Hoffmann, et al. (1997). "Stage-specific expression of Dlx-5 during osteoblast differentiation: involvement in regulation of osteocalcin gene expression." Mol Endocrinol **11**(11): 1681-94.

Sambrook, J. and D. W. Russell (2001). Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press.

Shipley, N. (2005). Role of the I56i Enhancer Sequence in the Expression of the Dlx Gene Family. Biology. Ottawa, University of Ottawa. **B. Sc., Biology (Honours, Co-op):** 27.

Simon, H. H., S. Thuret, et al. (2004). "Midbrain dopaminergic neurons: control of their cell fate by the engrailed transcription factors." Cell Tissue Res **318**(1): 53-61.

Solomon, K. S. and A. Fritz (2002). "Concerted action of two dlx paralogs in sensory placode formation." Development **129**(13): 3127-36.

Stock, D. W., D. L. Ellies, et al. (1996). "The evolution of the vertebrate Dlx gene family." Proc Natl Acad Sci U S A **93**(20): 10858-63.

Stuhmer, T., L. Puelles, et al. (2002). "Expression from a Dlx gene enhancer marks adult mouse cortical GABAergic neurons." Cereb Cortex **12**(1): 75-85.

Sumiyama, K., S. Q. Irvine, et al. (2003). "The role of gene duplication in the evolution and function of the vertebrate Dlx/distal-less bigene clusters." J Struct Funct Genomics **3**(1-4): 151-9.

Sumiyama, K. and F. H. Ruddle (2003). "Regulation of Dlx3 gene expression in visceral arches by evolutionarily conserved enhancer elements." Proc Natl Acad Sci U S A **100**(7): 4030-4.

Szucsik, J. C., D. P. Witte, et al. (1997). "Altered forebrain and hindbrain development in mice mutant for the Gsh-2 homeobox gene." Dev Biol **191**(2): 230-42.

Toresson, H., M. Parmar, et al. (2000). "Expression of Meis and Pbx genes and their protein products in the developing telencephalon: implications for regional differentiation." Mech Dev **94**(1-2): 183-7.

Vachon, G., B. Cohen, et al. (1992). "Homeotic genes of the Bithorax complex repress limb development in the abdomen of the Drosophila embryo through the target gene Distal-less." Cell **71**(3): 437-50.

Wang, W., P. Lo, et al. (2000). "Hmx: an evolutionary conserved homeobox gene family expressed in the developing nervous system in mice and Drosophila." Mech Dev **99**(1-2): 123-37.

Wang, W. and T. Lufkin (2005). "Hmx homeobox gene function in inner ear and nervous system cell-type specification and development." Exp Cell Res **306**(2): 373-9.

Waterston, R. H., K. Lindblad-Toh, et al. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-62.

Wilding, L. and M. Gannon (2004). "The role of pdx1 and HNF6 in proliferation and differentiation of endocrine precursors." Diabetes Metab Res Rev **20**(2): 114-23.

Yee, S. P. and P. W. Rigby (1993). "The regulation of myogenin gene expression during the embryonic development of the mouse." Genes Dev **7**(7A): 1277-89.

Zerucha, T. and M. Ekker (2000). "Distal-less-related homeobox genes of vertebrates: evolution, function, and regulation." Biochem Cell Biol **78**(5): 593-601.

Zerucha, T., T. Stuhmer, et al. (2000). "A highly conserved enhancer in the Dlx5/Dlx6 intergenic region is the site of cross-regulatory interactions between Dlx genes in the embryonic forebrain." J Neurosci **20**(2): 709-21.

Zhang, H., G. Hu, et al. (1997). "Heterodimerization of Msx and Dlx homeoproteins results in functional antagonism." Mol Cell Biol **17**(5): 2920-32.

Zhao, G. Q., X. Zhou, et al. (1993). "Cartilage homeoprotein 1, a homeoprotein selectively expressed in chondrocytes." Proc Natl Acad Sci U S A **90**(18): 8633-7.

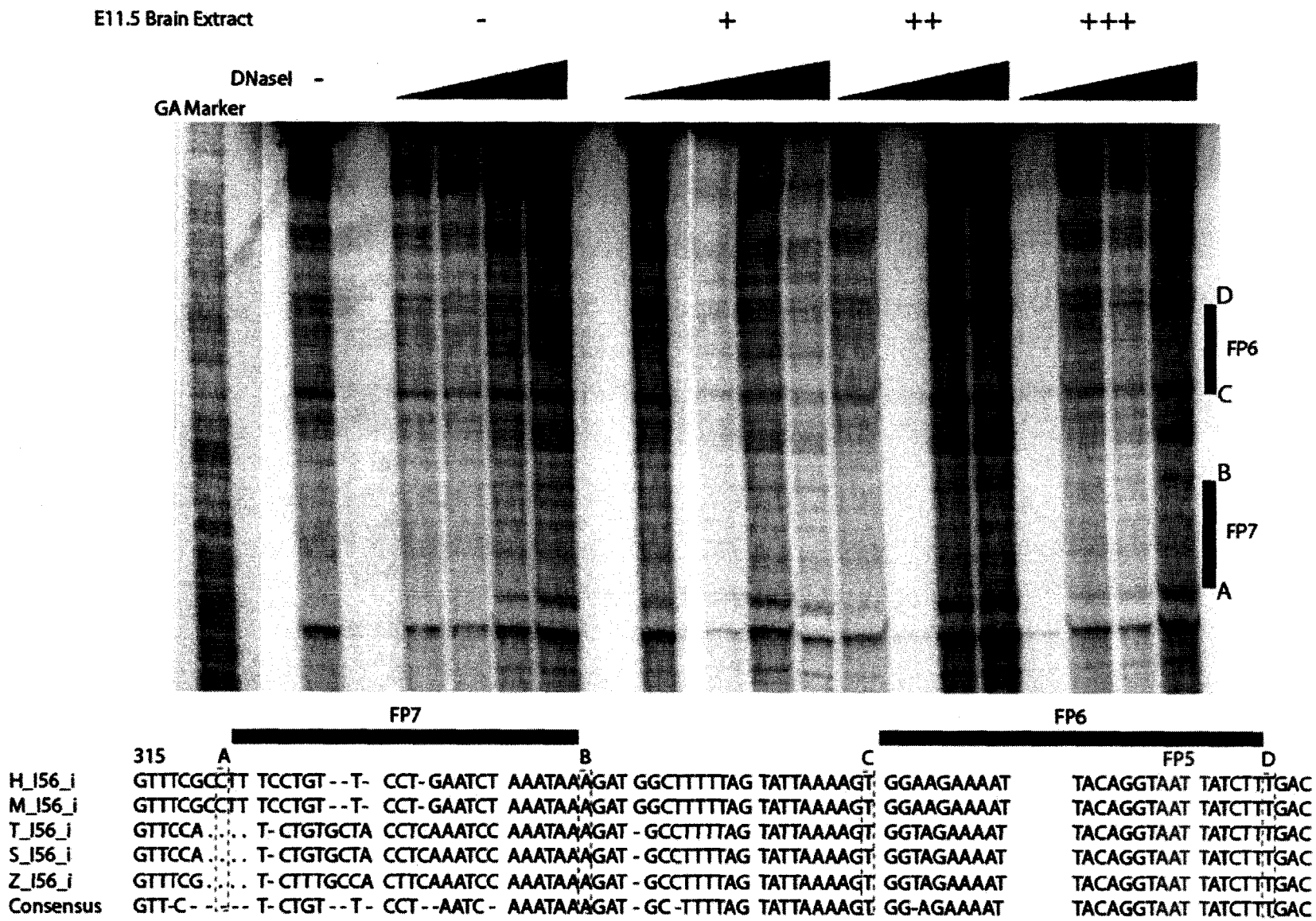
Zhou, Q. P., T. N. Le, et al. (2004). "Identification of a direct Dlx homeodomain target in the developing mouse forebrain and retina by optimization of chromatin immunoprecipitation." Nucleic Acids Res **32**(3): 884-92.

6. APPENDIX

Appendix 1.

Figure A.1. *DNaseI* footprinting confirms two putative binding sites within the I56i enhancer. Proteins from a nuclear extract of E11.5 mouse brain bind one fragment of the mouse I56i enhancer sequence at two sites, FP6 and FP7, previously identified by *DNase*I footprinting using E13.5 forebrain extracts. FP6 ranges from C to D and FP7 ranges from A to B. Increasing concentrations of nuclear extract are denoted by +, ++ and +++. Increasing concentrations of *DNase*I are denoted by the increasing slope. The previously identified TAAT core of the putative binding site within the FP5 protected region is shown in the sequence in blue. Data shown is representative of 2 repetitions.

The sequencing GA marker is shown on the left of each footprint. Thick lines denote the protected sites. The location of the footprint sites within the I56i enhancer sequence is shown below the footprint in a sequence alignment for five species [H = human, M = mouse, T = *Takifugu rubripes*, S = *Spheroides nephelus*, Z = zebrafish (*Danio rerio*)].


E11.5 Brain Extract
-
+
++
+++
DNaseI -
GA Marker
D
FP6
C
B
FP7
A
FP7
FP6
315
A
B
C
FP5
D
H_I56_i GTTTCGCCTT TCCTGT--T- CCT-GAATCT AAATAAAGAT GGCTTTTTAG TATTAAAAGT GGAAGAAAAT TACAGGTAAT TATCTTTGAC
M_I56_i GTTTCGCCTT TCCTGT--T- CCT-GAATCT AAATAAAGAT GGCTTTTTAG TATTAAAAGT GGAAGAAAAT TACAGGTAAT TATCTTTGAC
T_I56_i GTTCCA.... T-CTGTGCTA CCTCAAATCC AAATAAAGAT -GCCTTTTAG TATTAAAAGT GGTAGAAAAT TACAGGTAAT TATCTTTGAC
S_I56_i GTTCCA.... T-CTGTGCTA CCTCAAATCC AAATAAAGAT -GCCTTTTAG TATTAAAAGT GGTAGAAAAT TACAGGTAAT TATCTTTGAC
Z_I56_i GTTTCG.... T-CTTTGCCA CTTCAAATCC AAATAAAGAT -GCCTTTTAG TATTAAAAGT GGTAGAAAAT TACAGGTAAT TATCTTTGAC
Consensus GTT-C-.... T-CTGT--T- CCT--AATC- AAATAAAGAT -GC-TTTTAG TATTAAAAGT GG-AGAAAAT TACAGGTAAT TATCTTTGAC