INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Université d'Ottawa · University of Ottawa

EVOLUTION OF AUTO- AND CROSS-REGULATORY ELEMENTS OF MEMBERS OF THE *DISTAL-LESS*-RELATED FAMILY OF HOMEOBOX-CONTAINING GENES

Ted Zerucha

Thesis submitted to the University of Ottawa Department of Graduate Studies in partial fulfillment of the requirements for the degree of Doctor of Philosophy

University of Ottawa

Department of Cellular and Molecular Medicine

Anatomy and Neurobiology

Ottawa, Ontario, Canada

March, 1999

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

LIST OF ABBREVIATIONS

[α^{32}P]dATP	2'-deoxyadenosine 5'-triphosphate, α phosphorous labeled
A	Adenine
aa	amino acid
abd-a	abdominal-a
Abd-B	Abdominal-B
AEP	anterior entopeduncular area
AER	apical ectodermal ridge
AmphiDll	amphioxus Distal-less
ANT-C	Antennapedia complex
Antp	Antennapedia
bp	base pair
BSA	bovine serum albumin
BT	basal telencephalon
BX-C	Bithorax complex
C	Cytosine
CAT	Chloramphenicol Acetyl Transferase
cDNA	complementary deoxyribonucleic acid
CGE	caudal ganglionic eminence

CMV	cytomegalovirus
Cx	cortex
Dfd	Deformed
Dll	Distal-less
DNA	deoxyribonucleic acid
Dpp	decapentaplegic
DTT	Dithiothreitol
EC	embryonic carcinoma
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
en	engrailed
eve	even skipped
Exd	Extradenticle
ftz	fushi tarazu
G	Guanine
GFP	Green Fluorescent Protein

HEPES	4-(2-hydroxylethyl)-1-piperazine ethanesulfonic acid
HOM-C	Homeotic complex
h.p.f.	hours post fertilization
Hy	hypothalamus
lab	labial
LGE	lateral ganglionic eminence
LV	lateral ventricle
Md	mandibular component of the first branchial arch
MGE	medial ganglionic eminence
mRNA	messenger ribonucleic acid
MTG	myc-tag
Mx	maxillary component of the first branchial arch
MZ	mantle zone
OB	prospective olfactory bulb
OCA-B	Oct-1/Oct-2 B cell-specific coactivator
Oct-1	octamer transcription factor-1
Oct-2	octamer transcription factor-2
OD	optical density
OP	olfactory placode

pb	proboscipedia
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
Pit-1	pituitary transcription factor-1
PMSF	phenylmethylsulfonyl fluoride
POA	preoptic primordia
prd	paired
RNA	ribonucleic acid
Scr	sex combs reduced
Se	septum
so	sine oculus
SPV	supraoptic paraventricular area
SV40	simian virus 40
SVZ	subventricular zone
T	Thymine
TBE	Tris-borate/EDTA electrophoresis buffer
TBP	TATA binding protein
Tk	Thymidine kinase
Tris	tris(hydroxymethyl)aminomethane

Ubx	**Ultrabithorax**
VT	**ventral thalamus**
VZ	**ventricular zone**
wg	**wingless**
Xdll	*Xenopus* **distal-less**
zen	**zerknüllt**
zf	**zebrafish**

ACKNOWLEDGEMENTS

First and foremost I would like to thank my supervisor Marc Ekker for taking me into his lab, for providing funding during my tenure, for his patience, for being a mentor and for being a friend. I would also like to thank and acknowledge the people from the Ekker lab who were and are directly involved in this project: Gary Hatch for his work with the transgenic mice, DNA sequencing and for being my morning coffee buddy; Genny (Where's my Pepsi) Giroux for her involvement in the beginning of our ventures into mouse transgenics, her contributions to the cloning of the enhancers, DNA sequencing and for being a closet Tom Jones fan; Keon Park for the beautiful fish coronal sections in chapter 3; Qiaoming Long for the GFP fish in chapter 3; and Jennifer Miles for the sequencing of the entire zebrafish dlx3/dlx7 intergenic region.

Part of the work described in this thesis was done in collaboration with a number of labs that I would also like to acknowledge. Josh Schultz, presently at Tularik Inc. in San Francisco, was very helpful when we first began experimenting with transgenic mice as was Adrianna Gambarotta, who has continued doing our injections. Much of the analysis of the transgenic work described in chapter 3 was done with the assistance of John Rubenstein and his laboratory at the University of California San Francisco, in particular I would like to thank Thorsten Stühmer for his mouse coronal sections shown in chapter 3 and for many helpful e-mail discussions. I would also like to acknowledge Ken Weiss (Pennsylvania State University), David Stock (presently at the University of Colarado), Frank Ruddle (Yale University) and Webb Miller (Pennsylvania State University) for their contributions of the human and mouse Dlx3/Dlx7 intergenic

sequence and Webb Miller for the sequence alignment. The transgenic fly experiments shown in chapter 4 were done in collaboration with Paul Lasko's laboratory at McGill University. I would like to extend a special thank you to the Lasko lab, in particular Paul for kindly allowing me to come and work in his lab and his student Pascal Lachance for the friendly and patient crash course in fly husbandry. I would also like to acknowledge Jean-Pierre Muller for performing the RNA injections and in situ hybridizations shown in chapter 2.

A number of plasmids were kindly made available to me and I would like to thank Mark Featherstone (McGill University) for the *Pbx* expression plasmids, Martin Petkovich (Queens University) for the pTL plasmids, Serge Côté for the pWHIZ plasmid (and for being involved in giving the first shot to see if the zebrafish *dlx* enhancers may work in flies) and Gratien (Grass) Préfontaine for the pTL-MTG plasmid.

For their technical assistance I would like to thank: Lucille Joly for teaching me the calcium phosphate precipitation transfection method as well as for her friendship, the many 7:00 am bitch sessions, for letting most of my tapes get at least half way done before shutting them off, and for letting the cool, funky 70's one finish every time!; Ward Giffin for all the help with getting my shifts set up, for a lot of advice during the failed Dnase I footrpinting and ExoIII fiascos, for his friendship, for not killing me on any of our runs, for a whole lot of beer and wings and for the great fishing trip (yeah that last part was sarcastic); Grass for all the help when I first arrived with making protein, for his friendship and for being the best closest thing to a room mate I ever had (just remember to be careful travelling time without me to look for you!).

I would like to thank the past and present members of our lab family who have yet to be mentioned for providing a wonderful atmosphere to work in: Marie-Andrée Akimenko for her kindness and genuine love and excitement for science; Cris (no man, you're the king) Martin and Christof Nolte for their friendship and the, at times, absolute pandemonium, both of you guys really made a difference, thanks (also thank you Christof for coming out for the defence, that meant a lot, and Cris thanks for the many phone conversations and encouragement during the month leading up the defence); Deb Ellies for her friendship and for making my transition to living in a strange city an easy one; German Poleo for being German and for his friendship; Cricho Brown for his friendship although I think I could have definitely done with one less "Apocalypse Now" monologue; Liz Quint for her friendship, fuzzy duck, shut it etc and New Years slam dancing; Lynda Laforest for her friendship and constant smile; all the assorted summer (with special warm wishes for Stan Spacek), project and co-op students who came our way and in some cases probably taught as much or more than they learned; and last, but by no means least, Nathalie Chartrand, there are just no words, thank you for everything especially reinstilling my faith in people.

I would like to thank the Loeb Institute. I would also like to thank several people at the Loeb who have yet to be mentioned: Terry Reich for making me feel like part of the family, for all the runs, for the financial advice and for putting up with me living in his basement my last month in Ottawa (in regards to this last bit thank yous must also be extended to Nicole and Jeff Reich for putting up with the freak living under the stairs); Kursad Turksen for the great baseball analogy pep talk, and for lots of thoughts/advice on

life and maintaining cells; Marcin Boruk for being part of the long strange trip; and along with Ian (Lothar, no-carb) Laquian, Brian Hsu, Dave Rhodda, Dave Grandmaison, Simon Ginsberg, and Jo Savory I thank all of you for your friendship. Thank you also to Julianna Tomlinson for the totally excellent cheer. Finally at the Loeb I would like to thank Nuch Tanphaichitr for *hands* down the best part time job ever.

I would like to thank the members of my advisory committee, Marie-Andrée Akimenko, Max Hincke and Rob Haché. I would also like to thank the examiners for my comprehensive examination, Christine Pratt, Paul Albert, May Griffith and of course the chair Len Maler. I would like to extend a special thank you to Bill Staines for essentially being a fourth member of my advisory committee, for always being available to talk, for his kindness and for always willing to do what he can to help. Also, thank you to Pierre Fortier for all his work trying to organize the grad students in our department, for his kindness and his constant availability. I would also like to thank Shannon Goodwin for all of her help in getting this thesis in. Thank you to the members of my thesis examining committee, Bill Staines, Dave Parry, Dave Picketts, Pat Krone and the chair Vas Mezl. Briefly I would also like to acknowledge the contributions to sound mind made by: Sedykh; the miracle of life pigeons from the Parkdale underpass at the 417; Mr. and Mrs. Long for all the poutine; Todd at Pooley's for the 18 year old MacAllan single malt (I raise my glass to the memory of Pooley's, which tragically burned down April 10, a very sad day); and the guy at the Carleton who remembers I like Blue.

Finally I would like to thank my parents, family and friends from back in Winnipeg for their support and encouragement. In particular I must thank my mother and father for their never ending support especially in the weeks leading up to the defence. I would like to thank my brother Tony for his support and huge heart, my grandmother and my aunt and uncle Marilyn and Nigel Lilley for their support and encouragement. Thank you also to Suke and Lisa Claire and Austin Harvey and Elizabeth Pate for all of their support and encouragement particularly towards the end. A special thank you to Austin for being there in the clutch yet again. It was strangely appropriate having at least one medicine session for this degree.

ABSTRACT

The Vertebrate *Dlx* gene family consists of at least three linked pairs of convergently transcribed genes. The linked *Dlx* genes demonstrate a striking degree of overlap in expression patterns during development that is coincident with their genomic organization and is relatively well conserved between teleost fish, such as the zebrafish, and mammals. Two explanations for this overlap of expression are that there are cross-regulatory interactions between the *Dlx* genes or, in the case of the paired *Dlx* genes, that they share *cis*-acting regulatory elements. To test this second hypothesis, we compared the homologous intergenic regions of the paired *Dlx* genes between mouse and zebrafish. We have identified elements in the intergenic regions of all the paired *Dlx* genes that are well conserved between zebrafish and mice and that may, at least partially, account for the overlap in expression patterns of their neighboring *Dlx* genes by acting as shared enhancer elements. At least two of these elements are able to direct specific expression in transgenic mice and, remarkably, also in analogous anatomical regions in transgenic flies. Two of the elements identified are also the sites of regulation by the Dlx proteins themselves. An additional site of regulatory interactions between members of the *Dlx* family has been identified in the region upstream of the zebrafish *dlx4* gene which seems to be dependant on expression of *dlx3*. These results indicate that interactions between members of the *Dlx* family are part of their normal function during development and that there has been some degree of conservation of the elements directing expression of the *Dlx* genes during metazoan evolution. These results also suggest a mechanism by which duplicated genes may be stabily maintained during evolution as well as a possible explanation for the maintenance of the *Dlx* genes as linked pairs in the genome.

1. INTRODUCTION

"*The microcosm: ontogeny. The macrocosm: cosmic history, human history, organic development*" (Gould, 1977). This statement from one of the opening chapters of Gould's Ontogeny and Phylogeny simply, and poetically, describes the far-reaching importance of understanding development. Studying the mechanisms of how organisms related to ourselves develop has had and continues to have far-reaching implications in the medical and life sciences. At the same time, the realization that many of these mechanisms are remarkably well-conserved in diverse phyla has added a completely new dimension to the study of evolution.

In the last 10-20 years many of the molecular factors involved in metazoan development have been identified. In embryos, transcription factors regulate the patterning of the embryo as well as the determination of cell fates. This regulation may be based on specific interactions between transcription factors and their target genes and/or other transcription factors. A large number of transcription factors are sequence-specific, DNA-binding proteins. Many different families of DNA-binding transcription factors have been identified in eukaryotic organisms, one example consists of the factors coded for by the super family of homeobox genes.

The *Dlx* genes make up one of the homeobox gene families that belong to the super family of homeobox genes. The most well characterized members of the homeobox gene super family are the vertebrate *Hox* genes and their *Drosophila*

homologues the homeotic cluster (*Hom-C*) genes. The *Dlx* genes share a specific feature with the *Hox/Hom-C* genes in that they are clustered in the genome. This feature is not common to all homeobox genes, and while the significance of the clustering of the *Dlx* genes is only beginning to be understood, the significance of the clustering of the vertebrate *Hox* genes is relatively better understood. This significance is largely based upon the relationship between the clustering of the *Hox* genes and their expression as well as regulation. Because of this, the mechanisms of the relationship between gene clustering and their expression will be discussed, in terms of the *Hox* genes, in order to present ideas that will be important for understanding the relationship between genomic organization and regulation of expression of the *Dlx* genes.

Leading up to this, the current state of knowledge of the evolution of the *Hox* and *Hom-C* genes will be discussed. This will serve to illustrate not only how the modern *Hox* genes arose but will also serve to introduce the importance of gene and larger chromosomal duplication events in the evolution of clustered genes in particular and gene families in general. In addition, this discussion on the evolution of the *Hox* genes will illustrate the usefulness of examining homologous genes in diverse species to understand how they have evolved. I will then bring together the topics of *Hox* gene regulation and their evolution in a brief discussion of how the mechanisms that control gene expression may be conserved in evolutionarily diverse organisms.

Following this introduction to the evolution and significance of the organization of the *Hox* genes, the actual function of the products of homeobox-containing genes will

be discussed. This section will include an examination of the ability of homeodomain proteins to bind DNA, mediated by their homeodomains, and to regulate gene transcription.

At this point, the evolution and relationship between genomic organization and regulation of expression of the clustered *Hox* genes will have been discussed, as will have the general function of homeodomain proteins, and the *Distal-less*-related family of homeobox-containing genes will be introduced. I will begin by introducing the first member of this family that was identified, the single *Drosophila Distal-less* (*Dll*) gene. Relative to the other members of this family, the *Drosophila Dll* gene is the most well characterized, in terms of physiological role and developmental regulation. Thus it provides a useful starting point to begin to understand the role this gene family plays in development as well as the factors that regulate its expression. Because of the tendency for regulatory cascades to be re-used in nature as organisms evolve, the understanding of the more simplistic setting in which the single *Drosophila Dll* acts can likely provide hints on the regulation and function of the vertebrate *Dlx* genes.

At this point the vertebrate *Dlx* genes will begin to be discussed. The *Dlx* genes from different vertebrate species will be briefly introduced as was done for the *Hox* genes. The genomic organization and number of vertebrate *Dlx* genes as well as their invertebrate homologues will be incorporated into a reconstruction of the likely evolutionary history of this gene family. The expression of the vertebrate *Dlx* genes will be described and then their proposed roles at the molecular level and at the level of the

entire organism as deduced from functional inactivation studies. This will complete the description of the current state of knowledge pertaining to my research project and the introduction will be completed with the statement of the problem addressed by my project.

The Homeotic Genes

The homeodomain, or the genetic component which encodes it, the homeobox, are named for the homeotic fly mutations which facilitated the discovery of this element. The first *Drosophila* homeotic mutant was discovered by Calvin and Bridges in 1915 (Bridges and Morgan, 1923) although the term homeotic had been coined several years before by the famous British biologist William Bateson (Bateson, 1894). Bateson used homeotic to describe a malformation that substitutes the pattern of one region to that of another and he was referring to systems of reiterated elements in organisms such as the petals of a flower, the segments of an arthropod or the vertebrae of vertebrates. Bateson specifically used homeotic to describe a malformation in which there has not "*merely been a change, but that something has been changed into the likeness of something else*". This was consistent with the malformations observed in the *bithorax* mutant described by Bridges and Morgan. This mutation has, as a phenotype, a partial conversion of the haltere (a small balancing appendage on the thoracic segment directly posterior to that of the wing) to a wing. The second *Drosophila* homeotic mutant identified was similar but where the *bithorax* mutation involved an anteriorization (a posterior segment taking on

characteristics of a more anterior segment), the *antennapedia* mutant (Gehring, 1967) involved a posteriorization. The antennae of the adult fly are transformed into legs.

The subsequent cloning of the *antennapedia* and *bithorax* loci revealed that each mutation mapped to a cluster of genes, the *antennapedia* complex (*ANT-C*) and *bithorax* complex (*BX-C*) (Lewis, 1978). Both of these clusters are physically located close to one another in the *Drosophila* genome and the *BX-C* contains 3 genes and the *ANT-C*, 5 genes. Interestingly, mutants in either of these loci map in order along the chromosome to reflect the effects of the mutations on the antero-posterior axis of the fly. Thus, the left-most (3') mutations in a cluster affect the most anterior body parts. This is also consistent with the expression patterns of these genes in that they are expressed spatially and temporally in the order in which they sit in their clusters. Genes at the 3' end of the clusters are expressed first and more anteriorly than the genes more 5' in the clusters. Because these genes seem to have similar functions, although in different places, and are localized to regions of the genome associated with gene duplication events, Lewis (1978) proposed that the genes in these clusters may have evolved by duplication and divergence from an ancestral gene. Such a mechanism of duplication then divergence had also been postulated previously by Bridges and Morgan (1923), and such a diversification of the homeotic genes could have contributed to the evolution of insects with their varied body segments from ancestors having simpler repeating patterns of body segments.

Based on this proposal, Lewis (1978) predicted that if the genes in these complexes did share a common ancestor and maintained a similar function, the genes within the complexes should have similar structures. The cloning of the genes within these complexes revealed this prediction to be true. Two independent groups identified a highly conserved 180 bp element within several of the coding regions of the genes of the *HOM-C* which consists of the *ANT-C* and *BX-C* (McGinnis et al., 1984; Scott and Weiner, 1984). This element was named the homeobox and the protein domain it encodes, the homeodomain. Subsequently, genes containing a homeobox were identified not only outside the *HOM-C* in *Drosophila*, but in every eukaryotic species in which they were searched for. The most recent comprehensive survey of homeobox genes lists over 200 different ones which have been identified (Duboule, 1994) in phyla ranging from yeasts to plants to vertebrates. In addition, homologues of the *Drosophila HOM-C* clusters have now been identified in virtually all metazoans.

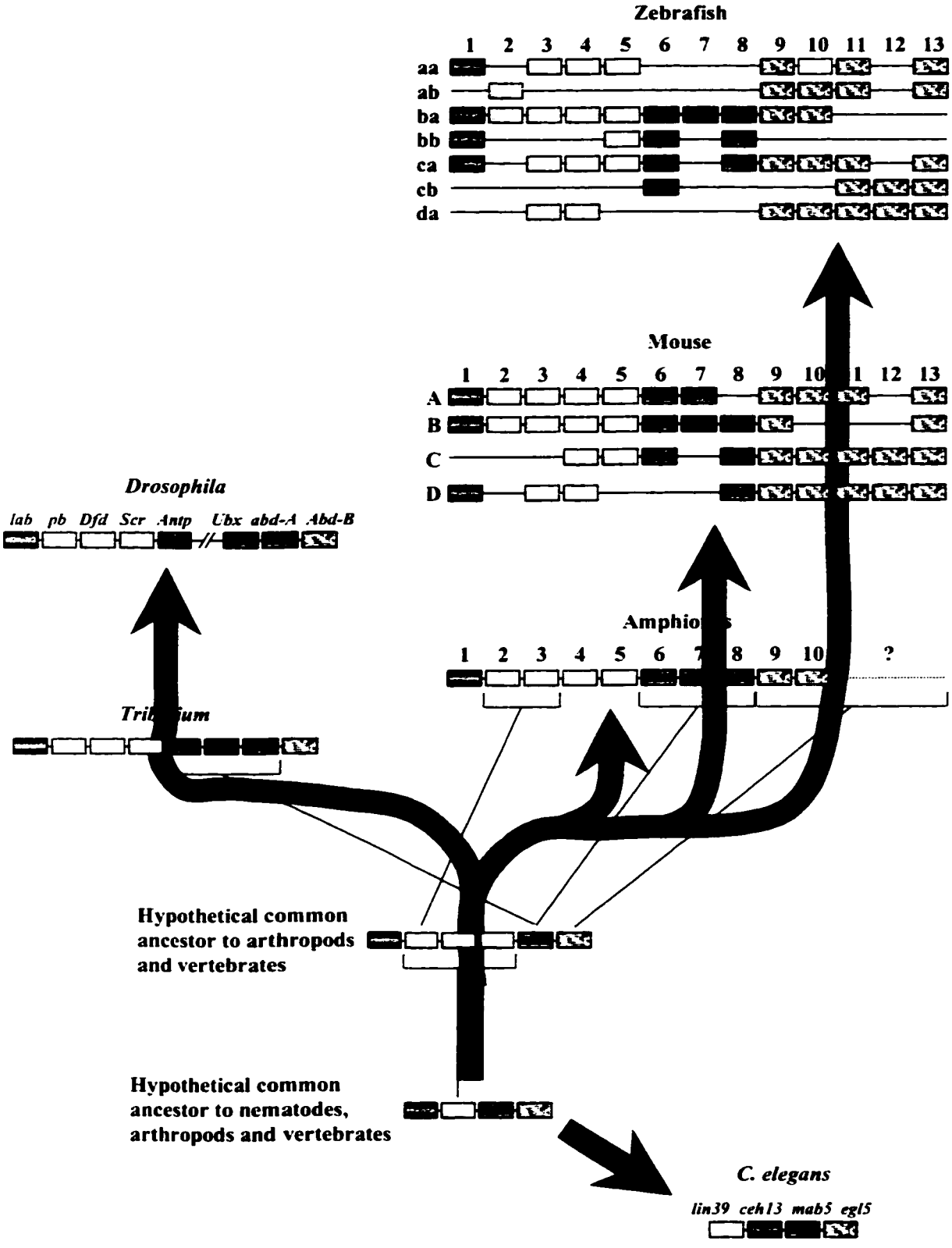
Evolution Of The Homeotic Genes

The vertebrate homologues of the *HOM-C* genes have been named the *Hox* genes. While in *Drosophila* there is only one set of *HOM-C* genes, in vertebrates as many as 7 sets of *Hox* genes have been identified. In arthropods, such as insects, 8 *Hom-C* genes have been identified and while these genes are separated into 2 clusters in *Drosophila*, in other insects such as the beetle *Tribolium*, the 8 genes are organized into one cluster. This suggests that the separation of the cluster in *Drosophila* has occurred relatively recently in its evolution. The nematode worm *Caenorhabditis elegans* has 4 *Hox* gene

homologues (Kenyon, 1994). The cephalochordate amphioxus, the closest invertebrate relative of vertebrates has at least 10 *Hox* genes organized into a single cluster (Garcia-Fernandez and Holland, 1994). The mouse, representative of mammals, has 39 *Hox* genes organized into 4 clusters of 13 paralagous groups. Finally, and in contrast to the common view that number of *Hox* clusters is a reflection of organismal complexity, in the teleost zebrafish 47 *hox* genes have been identified in 7 clusters of 13 paralagous groups (Amores et al., 1998).

By examining the organization and number of *Hox* genes and their homologues in diverse phyla a picture of the evolution of this gene family can be inferred (Fig. 1.1). Homologies between *Hox* genes are based on amino acid sequence similarities of the homeodomains which they encode. The proposed evolution of the *Hox* family involves gene duplication events, duplications of larger chromosomal regions and the loss of genes. A hypothetical ancestor to modern nematodes, arthropods, and vertebrates had a single *Hox* cluster of 4 genes (Kenyon, 1994)(Fig. 1.1). In the lineage that would give rise to modern nematodes, the positions of the first and second genes in the cluster are exchanged. In the transition from this hypothetical ancestor to the common ancestor of arthropods and vertebrates, the second gene in the cluster goes through at least 2 duplications resulting in a cluster of 6 genes. In the lineage that would give rise to modern insects, the second most 5' gene in the cluster undergoes a series of 3 duplications to give rise to the ancestral *Antp*, *Ubx* and *abd-A* genes as seen in modern insects. This last series of duplications results in a cluster of 8 genes which is observed in modern insects and which is schematized for the flour beetle *Tribolium*. In the lineage

Figure 1.1. A proposed model for the evolution of *Hox* gene complexity as schematised through what is known about *Hox* gene number and organization in a number of modern metazoan species. Genes are shown as boxes and likely homology to one of the four proposed ancestral *Hox* gene ancestors is represented by conserved shading within each box. The non-filled box in the position of zebrafish *hoxaa10* indicates a pseudogene and spaces indicate the absence of that gene. In the cases of vertebrates and amphioxus, the numbering above each gene represents the paralagous group to which it belongs and the cluster name is indicated by letters to the left of each cluster. Thus the top, left-most gene shown in the murine *Hox* clusters is *Hoxa1*. In the case of the Drosophila *HOM-C*, the genes are labelled with the standard abbreviations according to the nomenclature. The branched arrows indicate branching of metazoan lineages and reflect the evolution of the *Hox* clusters shown for each species, but are not reflective of the historical timing of these events. The arrows indicate a net increase in complexity based on number of *Hox* genes and clusters and are in no way meant to suggest fitness.


Zebrafish
1 2 3 4 5 6 7 8 9 10 11 12 13
aa
ab
ba
bb
ca
cb
da
Mouse
1 2 3 4 5 6 7 8 9 10 11 12 13
A
B
C
D
Drosophila
lab pb Dfd Scr Antp Ubx abd-A Abd-B
Amphio
1 2 3 4 5 6 7 8 9 10 ?
Tri ium
Hypothetical common ancestor to arthropods and vertebrates
Hypothetical common ancestor to nematodes, arthropods and vertebrates
C. elegans
lin39 ceh13 mab5 egl5

that would ultimately give rise to modern *Drosophila*, a genomic event resulted in the split of the cluster.

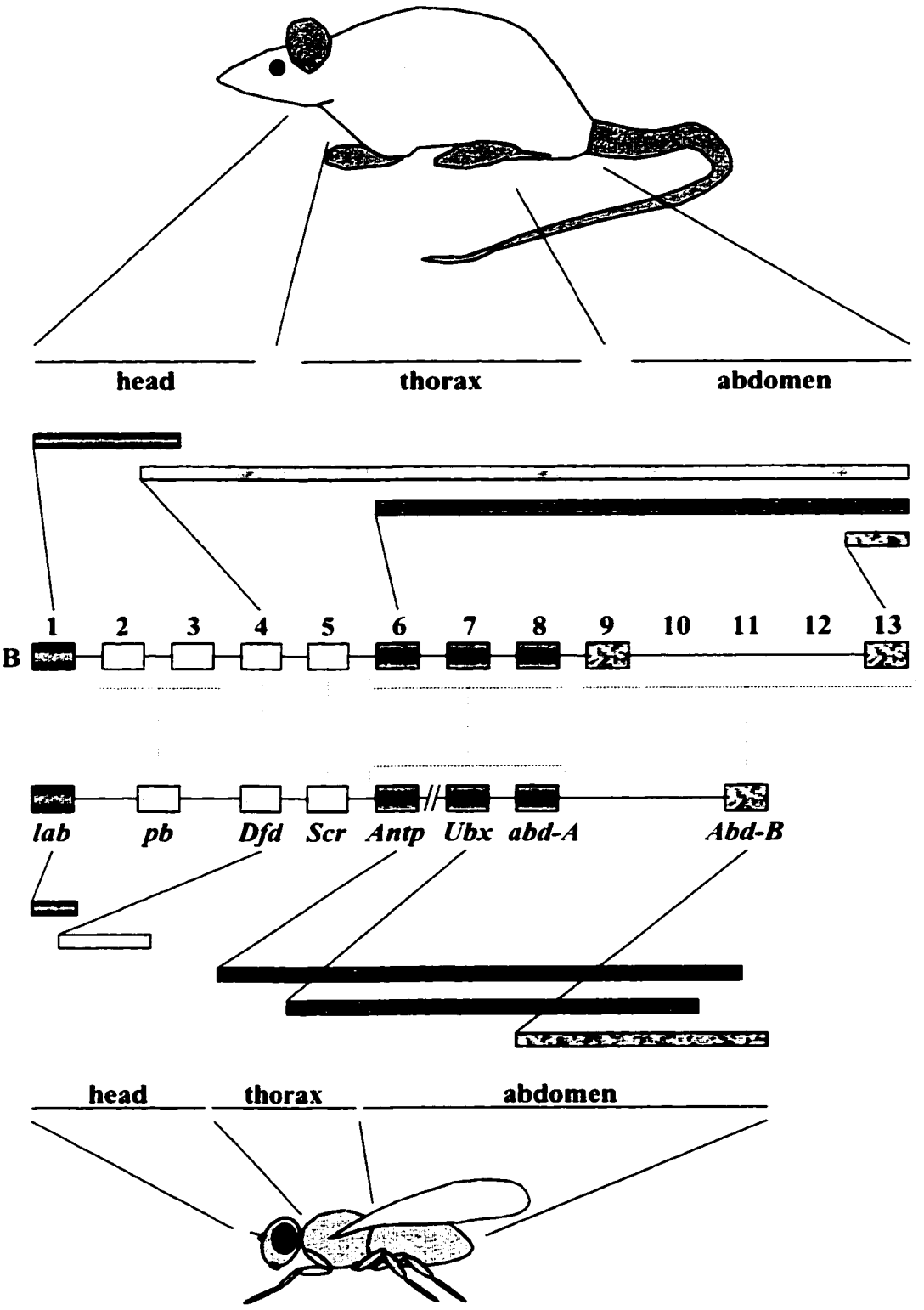
In the evolution of vertebrates, it seems that the ancestral *Hox* cluster went through a series of gene duplication events resulting in a primitive vertebrate state of 1 cluster of 13 genes. It is generally believed that the cephalochordate amphioxus represents the closest invertebrate relative of the vertebrates and consistent with its position as representative of a primitive vertebrate, it contains one *Hox* cluster of at least 10 genes (Garcia-Fernandez and Holland, 1994). These 10 genes belong to each of the first (1-10) paralagous groups of *Hox* genes observed in vertebrates. The presence or absence of members of the paralagous groups 11-13 remains to be determined. Regardless of the presence of a member of each of the vertebrate paralagous groups in amphioxus, it is apparent that the common ancestor of vertebrates did have at least one *Hox* cluster of 13 genes. Thus, until the presence or absence of additional *Hox* genes in the amphioxus *Hox* cluster is determined, it is unclear if amphioxus arose from the same common ancestor as modern vertebrates or if the lineage which would give rise to it underwent further gene duplications to yield the ancestral vertebrate state. It is believed that this ancestral cluster of 13 genes underwent a duplication resulting in 2 clusters, a proto-(AB) cluster and a proto-(CD) cluster (Amores et al., 1998; Kappen and Ruddle, 1993). Following this duplication it is presumed that the group 12 paralogue from the AB cluster and the group 2 and 7 paralogues from the CD cluster were lost since no paralogues from these groups have been identified (group 12 in clusters A or B or their derivatives, nor groups 2 and 7 in clusters C or D or their derivatives). Another

duplication resulted in the 4 *Hox* clusters as seen in mammals followed by more gene loss, specifically: *Hoxa8*; *Hoxb10* and *b11*; *Hoxc1* and *c3*; and *Hoxd5* and *d6*. After the branching of the lineage which would give rise to mammals, apparently the ancestor of ray-finned fish including zebrafish and the pufferfish *Fugu* underwent a polyploidization resulting in 8 *Hox* clusters. This was followed by the loss of a complete cluster from the zebrafish lineage, of 4 clusters from the *Fugu* lineage as well as individual genes from the remaining clusters (Amores et al., 1998). This model for the evolution of the *Hox* genes thus provides an excellent example illustrating the importance of small scale gene duplication events as well as larger chromosomal and genome duplication events in the evolution of a family of genes.

Homeotic Gene Expression: Colinearity And Posterior Prevalence

In vertebrates, the expression pattern of the *Hox* genes along the antero-posterior axis of the developing embryo is consistent with that already described for *Drosophila* and is schematized in Fig. 1.2. Genes at the 3' end of the cluster are expressed first and more anteriorly (e.g. the hindbrain) than genes more 5' in the cluster which are expressed more caudally. In addition, mutations in the *Hox* clusters are able to produce homeotic phenotypes similar in nature to that observed in *Drosophila*. For example mice which have had both copies of the coding sequence for the *Hoxc8* gene removed, before dying shortly after birth, exhibit transformations consistent with a homeotic anteriorization (Le Mouellic et al., 1992). Specifically, the eighth pair of ribs are directly attached to the sternum and an extra fourteenth pair of ribs appear on the first lumbar vertebra. Both of

Figure 1.2. A schematization illustrating the conservation of colinear expression patterns of the *Hox* genes in vertebrates, as represented by the mouse, and insects as shown for *Drosophila*. The genes are labelled as in figure 1.1 with the addition that shared common ancestry is indicated by dashed lines. Thus *Dfd* and *Hoxb4* are likely homologous, and while *Hoxb6-8* likely share a common ancestor with *Antp*, *Ubx* and *abd-A* none of these six genes are true paralogues. The approximate expression domains along the antero-posterior axis are represented by bars and are shown only for selected genes to illustrate the colinearity of *Hox* and *HOM-C* expression.


head
thorax
abdomen
1
2
3
4
5
6
7
8
9
10
11
12
13
B
lab
pb
Dfd
Scr
Antp
Ubx
abd-A
Abd-B
head
thorax
abdomen

these transformations are consistent with posterior vertebrae assuming characteristics of more anterior ones. A generalization to explain these phenotypes in both *Drosophila* and vertebrates has been developed and is commonly referred to as posterior prevalence. This generalization, also known as posterior dominance suggests that the more posteriorly expressed *Hox* genes have the tendency to inhibit the action of *Hox* genes expressed anterior to them. Thus, changes in *Hox* gene expression generally affect the more anterior domains of their expression, leaving the posterior structures within their domains of expression virtually unscathed. Anteriorizations are also observed in mice with *Hox* mutations. One such mutant is induced by the ectopic expression of *Hoxa7*. These mice also die shortly after birth and exhibit multiple craniofacial abnormalities (Balling et al., 1989; Kessel et al., 1990). One of the malformations described included the addition of an extra vertebra anterior to the first cervical vertebra, the atlas. The difference then, between a posteriorization and anteriorization is in the type of mutation. Anteriorizations are generally produced by gain of function mutations, such as that described for *Hoxa7* above or previously for the *Drosophila Antp* mutant. In these types of mutations, a gene's domain of expression is ectopically extended to a more anterior domain than normal, thus inhibiting the actions of the endogenous anterior-acting *Hox*, or *HOM-C*, genes. Posteriorizations, conversely, are generally produced by loss of function mutations. Both of these explanations are consistent with posterior prevalence or dominance.

Transcriptional Regulation Of The Homeotic Genes

The striking similarities in the roles played by the *Hox/HOM-C* genes in metazoans as diverse as arthropods and vertebrates, as well as their apparent common origin, suggests that the role played by these genes during development must be of a relatively great fundamental importance in order to have been conserved over hundreds of millions of years of evolution. The conservation of the roles of these genes across the metazoan phyla is reinforced in experiments where vertebrate *Hox* genes are ectopically expressed in wild-type *Drosophila* embryos, resulting in phenotypes similar to that observed when the homologous *Drosophila* genes are similarly mis-expressed (Malicki et al., 1993; Malicki et al., 1990; McGinnis et al., 1990; Zhao et al., 1993). In addition, it has been shown that vertebrate *Hox* genes can functionally substitute for their *Drosophila* orthologues. The expression of chicken *Hoxb1* is able to, at least partially, rescue the wild-type phenotype in flies carrying a null mutation of the orthologous *labial* (*lab*) gene (Lutz et al., 1996) despite low sequence conservation outside of the homeodomain of the proteins encoded by the vertebrate and invertebrate gene.

In addition to this apparent conservation in function, the striking degree to which the clustered genomic organization and colinear patterns of expression of the *Hox* genes and their invertebrate cousins has been conserved, suggests there has been a high degree of selective pressure to maintain these characteristics during metazoan evolution. The debate on the explanation for the relationship between the genomic organization of the *Hox* genes and the coordination of their temporal and spatial patterns of expression has

yet to be resolved. However, studies examining transcriptional regulation within the *Hox* complexes have suggested at least a partial explanation for the clustering of these genes. The partial overlaps in expression patterns of neighboring *Hox* genes within a cluster hinted at the possibility that *Hox* genes might share *cis*-regulatory elements and research carried out in the laboratories of Denis Duboule and Robb Krumlauf has confirmed this hypothesis. Krumlauf's laboratory is principally interested in the *Hoxb* cluster and has identified a number of enhancer elements shared by neighboring *Hox* genes, as well as elements that specify expression for a single gene. For example a single enhancer element is shared by the mouse *Hoxb3* and *Hoxb4* genes, and directs expression to a domain caudal to the margin defined by the segmental boundary between rhombomeres 6 and 7 as shown by analysis of a transgenic reporter gene (Gould et al., 1997). However, while the pattern of expression directed by this enhancer is representative of the overlapping domain shared by *Hoxb3* and *Hoxb4*, the full *Hoxb3* expression pattern includes a region anterior to this domain, as is expected based on colinearity. To reconcile this, the authors also report the action of another enhancer which shifts *Hoxb3* expression rostrally, and that is not shared by *Hoxb4*. The position of the shared enhancer within the *Hox* cluster and its sequence are conserved between fish and mammals suggesting its evolutionary conserved role. In addition, the shared enhancer element is able to mediate transcriptional activation by the products of multiple *Hox* genes including *Hoxb4* and *Hoxd4*. This is of particular significance because it demonstrates not only the importance of enhancer sharing in the coordinated expression patterns of *Hox* genes, which further implies the importance of the clustering of the *Hox*

genes, but also the importance of auto and cross-regulatory interactions between *Hox* genes.

Research carried out in Duboule's laboratory, principally on the *Hoxd* cluster, has taken the concept of enhancer sharing beyond 2 linked genes sharing an intergenic enhancer. During development of the tetrapod limb, the posteriorly acting *Hox* genes (*Hoxd10-Hoxd13*), together with *Evx2* the 5' neighbor to *Hoxd13*, show nearly identical expression domains in the distal limb. Duboule's group transposed reporter transgenes within the posteriorly acting *Hoxd* complex and observed them to take on expression patterns similar to that of the posterior *Hoxd* genes in the developing limb (van der Hoeven et al., 1996). In addition, the replacement of the chromosomal region encompassing *Hoxd11-Hoxd13* with a reporter transgene revealed that early expression of the transgene within the limb bud was very similar to that observed for the endogenous genes that had been deleted, although later expression was not (Zákány and Duboule, 1996). The perturbations in this later expression may presumably be attributed to the loss of enhancers important for refining the expression of individual *Hox* genes, or to the loss of auto and cross-regulatory interactions resulting from the loss of several genes from the *Hoxd* cluster. The proposed explanation for this is the presence of a global enhancer located outside the *Hoxd* cluster, as defined by the limits of the coding regions of the posterior-most gene, which directs expression of the *Hoxd10-13* genes as well as *Evx2* in the developing limb. Further evidence to support this comes in the form of the murine X-ray induced *Ulnaless* mutation. This mutation has been proposed to be within the global enhancer as it maps to the vicinity of (but outside) the *Hoxd* cluster and leads to miss-

expression of the posterior *Hoxd* genes in the developing limb (Hérault et al., 1997; Peichel et al., 1997). It seems then, that the sharing of regulatory elements would provide a strong selective pressure to maintain the *Hox* genes as clustered, at least in vertebrates. It can not be ruled out however, that shared regulatory elements evolved after an ancestral *Hox* cluster and are thus perhaps an effect rather than cause of *Hox* organization. As Duboule states "*although the existence of such regulations would explain why genes would not 'like' to be separated from each other, they do not necessarily tell us why they were kept together in the first place*" (Duboule, 1998).

One piece of evidence that suggests that perhaps the sharing of enhancers was not instrumental in the conservation of the clustering of the *Hox* genes is that *Drosophila HOM-C* genes apparently do not share enhancer elements. Several vertebrate *Hox* enhancers have some activity in *Drosophila*, including the *Hoxb3/Hoxb4* shared enhancer described above, which has been shown to be able to be regulated by the products of not only the vertebrate *Hox* genes but also the *Drosophila HOM-C* genes (Gould et al., 1997). A short (30 bp) enhancer element has also been identified in several of the introns of the murine *Hox* genes including *Hoxa7* and *Hoxa4* and the *Drosophila Ubx* gene. These enhancers can be regulated by the products of both the vertebrate *Hox* genes and the *Drosophila HOM-C* genes (Haerry and Gehring, 1997; Haerry and Gehring, 1996; Keegan et al., 1997). Despite the ability of the enhancer elements, described above, to exhibit some activity in both flies and vertebrates, it seems that the *Drosophila HOM-C* transcription units, including the *cis*-regulatory elements that direct their expression, are insulated from their neighboring genes. The fact that the *HOM-C* is often interspersed

with non-homeotic genes, such as *fushi tarazu (ftz)* in *Drosophila*, that do not share expression patterns with the *HOM-C* genes implies this. Evidence supporting this insulation of transcriptional mechanisms comes in the form of mutations in the *HOM-C* which remove the boundaries separating two domains of regulation and thus allow neighboring genes to come under control of shared enhancers (Gyurkovics et al., 1990). This is also supported by observations on boundaries of chromatin domains within the *HOM-C* (Hagstrom et al., 1996; Karch et al., 1994; Zhou et al., 1996). The most striking evidence is implied by the actual division of the *ANTP-C* and *BX-C* in *Drosophila*. Intuitively, differences in the mechanisms of deployment of the vertebrate and invertebrate *Hox* genes should come as no great surprise. Insects such as *Drosophila* have unique stages of development in comparison to vertebrates. This includes stages such as the syncitial blastoderm and the hierarchy of developmentally early signaling events including the maternal coordinate genes, the gap genes, the pair rule genes and segment polarity genes which help organize the anterior-posterior axis prior to the expression of *HOM-C* genes. The establishment of the antero-posterior axis in vertebrates occurs differently and in nematodes, while their *Hox* homologues are expressed similarly along the antero-posterior axis, it is apparent that cell lineages are important to this mechanism as well. Thus it appears that certain mechanisms involved in *Hox* gene regulation have been conserved during metazoan evolution as evidenced by the ability to exchange enhancers between species as divergent as *Drosophila* and mice. It is also apparent however, that different metazoans have evolved unique pathways, at the molecular level, to deploy the *Hox* genes and their homologues. This is a common and important theme in the study of evolution. While mechanisms such as those

described above may be conserved, there is also a tendency for innovations to arise in the evolution of specific lineages. This allows organisms different avenues to pursue in evolving new mechanisms for evolutionary success.

Function Of Homeodomain-Containing Proteins: DNA-Binding And Regulating Transcription

To this point the function of the *Hox* genes during development has only been briefly touched upon. Based on mutant analysis, the importance of the *Hox* genes in vertebrate and invertebrate development has been demonstrated in phenotypic terms. The auto- and cross-regulatory interactions between members of the *Hox* genes has also been briefly discussed, implicating the products of the *Hox* genes as regulators of transcription. However, how the products of the *Hox* genes and homeobox-containing genes in general act at the molecular level has been largely ignored. From its first discovery, the sequence similarity of the homeodomain to the helix-turn helix motif of prokaryotic gene regulatory proteins had been noted (Laughon and Scott, 1984; Shepherd et al., 1984), as well as their similarity to the mating type proteins MATa1 and α2 in yeast (Shepherd et al., 1984) which are also regulators of transcription. These preliminary observations lead to the hypothesis that homeodomain proteins act as regulators of transcription and that the homeodomain represents the DNA-binding domain of these factors. This was verified as homeodomain proteins were demonstrated to bind DNA. Later, the homeodomain was shown to be sufficient for DNA-binding, the structure of the

homeodomain was determined and the ability of homeodomain proteins to regulate transcription was demonstrated.

The ability of homeodomain proteins to bind DNA was demonstrated by a number of groups in the 3-5 years following the initial reports describing the discovery of the homeobox (Beachy et al., 1988; Desplan et al., 1985; Desplan et al., 1988; Fainsod et al., 1986; Hoey and Levine, 1988; Laughon et al., 1988; Muller et al., 1988). One of the earlier studies identified a nucleotide sequence fragment upstream of the murine *Hoxa3* gene that was bound by nuclear extracts from mouse embryonic cells (Fainsod et al., 1986). To determine if Hoxa3 itself, that was likely present in those extracts (based on its embryonic expression pattern) could bind this region, it was recombinantly expressed in bacteria and subsequently shown to bind that fragment. A similar study assayed the ability of recombinantly expressed engrailed (en), a homeodomain-containing protein from *Drosophila* but not a member of the *HOM-C*, to bind a short repeated nucleotide sequence motif, TCAATTAAAT, found in an element known to be responsible for regulating *en* (Desplan et al., 1988). Full-length recombinant en was demonstrated to bind this sequence. In addition, by making use of deletions of the en protein, the authors also demonstrated that the homeodomain alone was sufficient for binding to this sequence. Interestingly the distantly related ftz homeodomain is also able to bind this sequence and both homeodomains were also shown to bind a repeat of the nucleotide sequence TAA. Further confirmation of the homeodomain as DNA-binding domain came when a purified polypeptide corresponding to the Antp homeodomain was

demonstrated to be sufficient for binding specific oligonucleotide sequences (Muller et al., 1988).

The research described above on the ability of en and ftz to bind the same en regulatory sequence and a TAA-repeat sequence, as well as demonstrating that the homeodomain represents a DNA-binding domain also hinted at a conserved DNA recognition site for homeodomains. Both nucleotide sequences, TCAATTAAAT and the TAA repeat, contain a common sequence of ATTA/TAAT. Subsequent studies utilizing binding site selection followed by determination of equilibrium constants and competition experiments, have demonstrated that homeodomains do seem to have a greater affinity for sequences containing this 4 nucleotide core element (Ekker et al., 1994; Ekker et al., 1991). One such report compares the DNA sequence preferences of homeodomain proteins encoded by four of the eight *Drosophila HOM-C* genes (Ekker et al., 1994). Of these four homeodomain proteins examined, three (Ubx, Antp, and Dfd) have a relatively high affinity for sequences containing an ATTA/TAAT core element. Exceptions to this trend do occur however, as the fourth homeodomain protein in this study, Abd-B, has a higher affinity for nucleotide sequences containing an ATAA /TTAT core element.

The obvious question that arises from the observations that homeodomains have similar nucleotide recognition sequences, concerns specificity. The issue of specificity in sequence recognition can best be addressed by first determining how the homeodomain interacts with DNA, beginning with the structure of the homeodomain. The three

dimensional structure of the homeodomain in solution was first determined for Antp by NMR spectroscopy (Qian et al., 1989). The Antp homeodomain consists of three α-helices (a fourth more flexible carboxy-terminal α-helix is also present although it is often referred to and considered as an extension of the third) and a flexible amino-terminal arm. The second and third α-helices form a helix-turn helix motif very similar to that observed in prokaryotic transcriptional regulatory proteins as had been suggested previously by Laughon and Scott (Laughon and Scott, 1984) and Shepherd and colleagues (Shepherd et al., 1984). The similarity of the second and third α-helices of the homeodomain to the prokaryotic helix-turn-helix motif was solidified further upon the solving of the homeodomain-DNA complex in solution, also by NMR (Otting et al., 1990). The homeodomain's extended third helix contacts the DNA in the major groove in a manner highly similar to that of the recognition helix of the helix-turn-helix motif. Unique to the homeodomain, additional base-specific contacts are made in the minor groove by the amino-terminal arm and the DNA backbone is contacted by the loop between the first and second helices as well as several amino acids in the amino-terminal region of the second helix. Highly similar structures of homeodomains bound to DNA have been determined for other distantly related homeo-proteins (Kissinger et al., 1990; Klemm et al., 1994; Wolberger et al., 1991; Wolberger et al., 1991) indicating the high degree of conservation of the mode of DNA binding by this motif. The core nucleotide element ATTA/TAAT that, as previously mentioned, is a preferred site recognized by the homeodomain represents a portion of the sequence in the major groove contacted by the third helix. Thus, intuitively, it would seem that specificity of DNA binding may be

conferred upon the homeodomain through its base specific contacts outside the major groove, possibly including those contacts made by the amino-terminal arm.

The importance of the amino-terminal arm in conferring binding site specificity outside the ATTA/TAAT core element has been reported by a number of groups. One of the first studies to examine this elegantly compared the amino-terminal arms of the murine Hoxd4 and Hoxa1 proteins (Phelan et al., 1994). The *Hoxd4* gene has an auto-regulatory function mediated through the binding of its protein product to a relatively well characterized element in its promoter. Hoxa1, though, is unable to bind and activate transcription through this same region. However, when the second and third amino acids of the Hoxa1 amino-terminal arm were converted to those in the corresponding positions of Hoxd4, the Hoxa1 mutant bound and activated transcription through the *Hoxd4* auto-regulatory motif. Based on what is known structurally of the homeodomain-DNA interaction, the amino acids at positions two and three of the amino-terminal arm contact the minor groove of the DNA double helix outside the ATTA/TAAT core sequence. Amino acids located further towards the carboxy-terminus such as at position five seem to impart some specificity within the core sequence as variations in that amino acid result in altered preferences for it. For example, the variation in core sequence preference between the *Drosophila* Ubx, Antp and Dfd proteins for ATTA/TAAT and Abd-B for TTAT/ATAA have been attributed to a variation at amino acid five in the amino terminal arm (Ekker et al., 1994). Similar alterations in core specificity which depend on amino acid variation at this position have also been demonstrated for Hox proteins (Phelan et al., 1997).

It appears that specificity can also be imparted by at least two different mechanisms. One general example by which this may be accomplished is homeodomain proteins binding DNA as part of a complex with other DNA-binding proteins. This type of mechanism would require more than one binding site, and would also possibly introduce the importance of spacing between binding sites. Another possible mechanism involves a protein containing more than one DNA binding domain which would require, then, a more complex binding site that would include a homeodomain core recognition sequence. A particularly well-characterized example of this latter possibility is the POU domain family of transcription factors.

The POU domain protein family was initially identified based on the observation that a number of known transcriptional regulatory proteins contained a conserved protein domain. This conserved POU domain was named based on the first letters from the four proteins first identified which contained this motif, the products of the mammalian *Pit-1*, *Oct-1* and *Oct-2* and the *C. elegans unc-86* genes (Bodner et al., 1988; Clerc et al., 1988; Finney et al., 1988; Herr et al., 1988; Ingraham et al., 1988; Muller et al., 1988; Scheidereit et al., 1988; Sturm et al., 1988; Sturm and Herr, 1988). The bipartite DNA-binding POU domain consists of two highly conserved regions separated by a variable linker. The ~75 amino acid POU-specific domain consists of four α-helices with the second and third helices forming a helix-turn-helix motif. The carboxy-terminal region of the protein forms a homeodomain. POU domain proteins bind similar nucleotide sequences which are commonly referred to as octamer motifs. The octamer motif and its variants consist of a sequence recognized by the POU domain and an adjacent A/T-rich

sequence recognized by the homeodomain. This family is able to bind DNA as monomers, homo-dimers and hetero-dimers. In addition phosphorylation has been demonstrated to be important for their function (Kapiloff et al., 1991; Segil et al., 1991a,b) as are interactions with other DNA-binding proteins such as nuclear receptors, basal transcription factors such as TATA-binding protein (TBP) and non-DNA-binding coactivators such as OCA-B.

The presence of additional DNA-binding domains in homeodomain proteins is not restricted to the POU domain family. Several members of the paired domain family of transcription factors which includes the products of the vertebrate *Pax* genes also contain a homeodomain in addition to the paired domain. The paired domain recognizes a palindromic nucleotide sequence with each half of the palindrome separated by several nucleotides, thus those members of the paired family which also contain a homeodomain essentially contain three DNA-binding motifs. Unfortunately, the characterization of homeodomain proteins such as those coded for by *Pax* genes or by any of the other homeobox genes already mentioned is not as comprehensive as that of the members of the POU domain family. It is relatively well understood how the POU domain proteins function at the molecular level to regulate transcription. However, the ability of homeodomain-containing proteins outside the POU domain family to directly activate and repress transcription in cultured cells has been demonstrated.

One of the first reports of proteins containing a homeodomain directly regulating transcription came from Jaynes and O'Farrell (Jaynes and O'Farrell, 1988). Their principle concern was that, up to that time, only circumstantial evidence suggested that homeodomain proteins were likely able to regulate transcription. Previous experiments had indicated the ability of Ubx to activate transcription through is own promoter while inhibiting the *Antp* promoter in tissue culture cells. However, these experiments could not address the possibility that the transcriptional effects might result through intermediates and not through direct interaction of the Ubx protein with these promoter elements. The authors placed previously identified homeodomain binding sites for the ftz proteins (Desplan et al., 1985) adjacent to minimal promoters and reporter genes. These reporter constructs were co-transfected into cultured cells together with a construct to express the *ftz* gene and an activation of several hundred fold was observed relative to the transfection of the reporter construct and no *ftz* effector construct. Further studies of a similar nature involving ftz as a transcriptional regulator revealed that co-expression with other homeodomain proteins including paired (prd) and/or zerknullt (zen) resulted in significant increases in transcriptional activation (Han et al., 1989). These results are consistent with a synergistic relationship between the products of these three genes. Interestingly the products of the *even-skipped* (*eve*) and *engrailed* (*en*) genes seemed to have an opposite effect as the expression of the products of these two genes in combination with the transcriptional activator products of the genes, *ftz*, *prd* or *zen* resulted in a repression of the previously identified activation. The important contributions of this work was that not only can the products of homeobox genes act as activators of transcription, but it seemed likely that these genes acting in combination was

important for their biological function. The complexity of the function of homeodomain proteins goes beyond them acting as either activators or repressors of transcription. For example the Ubx protein has been demonstrated to act as a repressor or activator of transcription on different promoters in co-transfection experiments (Krasnow et al., 1989).

One of the best understood examples of homeodomain proteins acting together is the interaction of the HOM-C proteins with Extradenticle (Exd) in flies and Hox proteins, with Pbx in vertebrates. These proteins cooperate with HOM-C/Hox proteins to interact with DNA thus imparting increased specificity as well as affinity for DNA-binding. The biological significance of this was made apparent by the demonstration that Pbx1 cooperates with Hoxb1 to auto-regulate *Hoxb1* expression (Pöpperl et al., 1995). Recent crystallography work has revealed that when homeodomain proteins from the Hox/HOM-C groups and the Pbx/Exd groups bind DNA cooperatively, both homeodomains make sequence-specific contacts such that contacts are made with the nucleotide bases and with the DNA helix backbone. These functional studies suggest the complex action of homeodomain proteins. Their ability to act in synergy, as well as to activate and/or repress transcription suggests they likely act together and that their action is largely dependant on not only the cellular context in which they are expressed (which would affect the complement of homeobox genes expressed) but also the *cis*-regulatory elements through which they act.

The *Distal-Less* Family Of Homeobox-Containing Genes

As more genes have been identified outside the *Hox* clusters, they have been grouped into families on the basis of sequence similarities of the proteins they encode and these genes have proven to be important to development as well. Families such as the *POU* domain and *paired* families have already been mentioned and another is the *Distal-less*-related family. The *Distal-less* (*Dll*) gene was first identified in *Drosophila* and homologues of it have subsequently been identified in virtually all metazoans. As was seen with *Hox* genes, vertebrates have multiple copies of *Distal-less* related (*Dlx*) genes which have presumably arisen as a result of gene and chromosomal duplications.

Distal-Less In *Drosophila*

Paralleling the discovery of the *Hom-C* clusters in *Drosophila*, the *Dll* gene was isolated by a chromosome walk to map the molecular lesions associated with several *Dll* mutant alleles (Cohen et al., 1989). The *Distal-less* mutations affect the segmented appendages of the adult as well as several of the larval sense organs (Cohen and Jurgens, 1989; Cohen, 1990; Cohen et al., 1989; O'Hara et al., 1993; Sato, 1984; Sunkel and Whittle, 1987). The adult structures affected include: the head appendages associated with the mouth such as the maxillary pulps, the labium, the labrum, the proboscis, a subset of the maxillary cirri; the antennae; and the legs. The larval structures affected include the labral, antennal, maxillary and labial sense organs of the head and the Keilin's sense organs of the thorax. The Keilin's larval sense organs are homologous to the distal

sensory apparatus of the legs of lower insects and are associated with developing adult appendages. These larval sense organs are rudimentary structures compared to the morphologically well developed limbs of simpler insects (such as caterpillars) in that they consist solely of the distal sensory apparatus. Sense organs not associated with developing limbs are unaffected in mutant embryos. In *Distal-less* mutants, the larval sense organs are deleted and the distal structures of the adult appendages are reduced in size and fused together. A number of different *Dll* mutations have been compared (Cohen and Jurgens, 1989; Cohen et al., 1989) and while there is some variability in the severity of the heterozygous mutations (homozygous mutations are embryonic lethal), discrete adult structures are never deleted. This indicates that *Dll* is involved in patterning the formation of the distal limb as a whole.

The appendages of arthropods are thought to have evolved from simple, non-segmented appendages present in annelid-like ancestors. Modern segmented appendages are believed to have evolved as a result of specializations added to the ancestral unsegmented appendage. Thus the proximal domain of modern segmented appendages is representative of an evolutionary ground state upon which modern specializations have been built. Based on its apparent role in distal but not proximal appendage structures, *Dll* has been proposed to be the gene which promotes the development of appendage structures above the developmental and evolutionary ground state (Cohen and Jurgens, 1989).

The expression domains of *Dll* during *Drosophila* development are consistent with their apparent function. Transcripts of *Dll* are first detected during cellularization of the blastoderm in a stripe that corresponds to the future maxillary and labial segments (Cohen, 1990). Prior to completion of cellularization, a second stripe corresponding to the antennal appendage primordium appears. Following shortly after the onset of germ-band extension, expression is detected in the labrum primordium. By the completion of germ band elongation, the maxillary and labial primordia expression domains have resolved, the labrum primordium has migrated to the anterior tip of the embryo and the leg primordia have appeared and are also expressing *Dll*.

Regulation Of *Dll* During *Drosophila* Development

Expression of *Dll* in the trunk is primarily dependant on the segment polarity genes *wingless* (*wg*) and *engrailed* (*en*) (Cohen, 1990). In *wg* mutant embryos, the labial appendage and leg primordia never develop. A similar phenotype is observed in *en* mutant embryos which also lack these structures and also the primordium of the maxillary appendage. Examining *Dll* expression in embryos mutant for either *wg* or *en* reveals a loss of expression in the primordia which would give rise to the deleted structures. The segment polarity genes are expressed as stripes in the embryo at the cellular blastoderm stage. The expression of *en* and *wg* define the anterior and posterior domains, respectively, of the parasegment boundaries which also coincide with the position of the leg primordia. In double labeling experiments *Dll* is expressed in cells surrounding those expressing *wg* in the leg primordia. This result is consistent with the

role of *wg* as a secreted signaling molecule. Comparison of *Dll* and *en* expression reveals an overlap of expression in the leg primordia, consistent with the role of en as a transcription factor (Cohen, 1990). Because of these expression patterns Cohen (1990) proposes that *Dll* is expressed in the limb primordia and its expression in the anterior compartment of the primordia is dependant on *wg* and in the posterior compartment upon *en*. In mutant embryos in which the domains of expression of *en* or *wg* are altered, the relationship between their expression and *Dll* is maintained; thus *Dll* is aberrantly expressed as well. A later study (Cohen et al., 1993) suggested that the positioning of *Dll* expression along the dorsal-ventral axis of the embryo is dependant on expression of *decapentaplegic* (*dpp*). Thus *Dll* is expressed in the thoracic segments of the embryo in the regions of overlap between the stripes of expression of *wg* which run along the dorsal/ventral axis and *dpp* which run along the antero-posterior axis. Interestingly, it seems that once expressed, *Dll* is able to maintain its expression by in turn activating *en*, *wg* and *dpp* (Gorfinkiel et al., 1997) as observed in experiments in which *Dll* was ectopically expressed in the embryo ultimately resulting in the development of ectopic ventral appendages.

Intuitively, the *HOM-C* genes are also excellent candidates for controlling *Dll* expression. Because establishment of a body with an antero-posterior axis with little doubt preceded the evolution of specializations such as appendages, it is likely that these specializations were added on to the already present antero-posterior body. Thus positioning of these specializations by the *HOM-C* genes would come as no surprise. Evidence for a relationship between the *HOM-C* genes and *Dll* can be inferred by loss of

function mutations in genes of the *BX-C* which result in the anteriorization phenotype of abdominal legs (Lewis, 1978). This suggests that one of the roles of the *BX-C* is to repress limb development in the abdominal segments of the *Drosophila* embryo perhaps through repressing *Dll* expression.

In fact, enhancer elements have been identified which direct *Dll* expression during development and which are under control of several of the *HOM-C* genes. The coding region of the *Dll* gene is encoded in 7 exons spanning approximately 20 kb of the genome and the homeobox is divided by an intron (Vachon et al., 1992). A number of *cis*-regulatory elements have been identified in the 15 kb upstream of the coding region, in the 10 kb downstream region and in the third intron which divides the homeobox (O'Hara et al., 1993; Vachon et al., 1992). These enhancer elements were identified by coupling fragments of genomic DNA encompassing the *Dll* locus with a basal promoter and the *LacZ* gene, transforming flies with this series of reporter constructs, and examining transgenic embryos for β-Galactosidase (Vachon et al., 1992). A number of elements were identified which have the ability to recapitulate certain domains of endogenous *Dll* expression. Two enhancers upstream of *Dll* , 215 and 304 direct expression in the developing legs (215 also directs expression in the developing head regions including the labium, maxilla, antennae and labrum), and another named ETD6, located downstream of *Dll* directs expression to the cells which will become the maxillary cirri.

The *Dll*-304 enhancer directed transgene is coexpressed with endogenous *Dll* in the thoracic leg primordia of the germband extended embryo whereas the expression driven by the *Dll*-215 enhancer is in only a small subset of the *Dll* expressing cells in the same regions. In embryos mutant for the *BX-C* genes, however, the expression domains of both *Dll* and the transgene directed by *Dll*-304 are extended posteriorly to include the homeotic leg primordia in the abdominal segments. This strongly suggests that *Dll*-304 is the site of transcriptional repression by the products of members of the *BX-C*. This was confirmed in gain of function *BX-C* mutants which have anteriorly extended domains of expression of genes of this complex, and in which *Dll* as well as transgene expression was repressed. Further Ubx and abd-A proteins bind the *Dll*-304 element.

The ETD6 element unlike *Dll*-304 appears to be the site of transcriptional activation by at least one of the homeotic gene products. Expression of a transgene driven by ETD6 is observed in the ventral-lateral maxillary segment (O'Hara et al., 1993), which is also encompassed within the expression domain of the *ANT-C* gene *Dfd*. In *Dfd* loss of function mutant embryos, expression conferred by ETD6 is completely abolished. In embryos in which *Dfd* is ectopically expressed however, expression directed by ETD6 extends posteriorly from its wild type expression domain, although it is not expressed in all cells expressing *Dfd*. The authors do not address this latter problem although it seems likely that one possible explanation for these observations is, that while ETD6 directed expression is dependant on Dfd, other co-factors present in the cells which will give rise to the maxillary cirri are also required for optimal expression.

It seems then that the general role of *Dll* in insect development is to elevate structures above an evolutionary ground state essentially consisting of the body wall. The *Dll* gene functions in development of virtually all of the adult appendages in *Drosophila* and appears to be regulated by complex patterns of extracellular factors and other transcription factors, a complexity which is just beginning to be understood. Similar roles for *Dll* have been observed in other arthropods including other insects and crustaceans (Panganiban et al., 1995). The expression patterns of *Dll* in crustaceans is especially interesting as crustaceans have abdominal limbs (as opposed to thoracic limbs for insects). In the developing abdominal limbs *Dll* is expressed despite expression of the *BX-C* genes suggesting the repression of *Dll* expression by these homeotic genes is a relatively recent innovation in the evolution of modern insects (Palopoli and Patel, 1998). In addition to altered *Dll* expression leading to variation in the evolution of arthropods, other work has demonstrated that it has also been recruited, or exapted, to perform new functions such as in patterning the eyespots of butterflies (Carroll et al., 1994).

The Vertebrate *Distal-Less* Genes

The exaptation of genes to evolve new functions, such as the role of *Dll* in patterning the eyespots of butterfly wings, is an important component of modern reductionist evolutionary thought. Exaptations are often associated with gene duplication events. For example a gene is duplicated resulting in 2 genes capable of performing the same function. This redundancy now allows the mechanisms of evolution to modify 1 of the genes while the other maintains its normal functions. Over time, selection acting

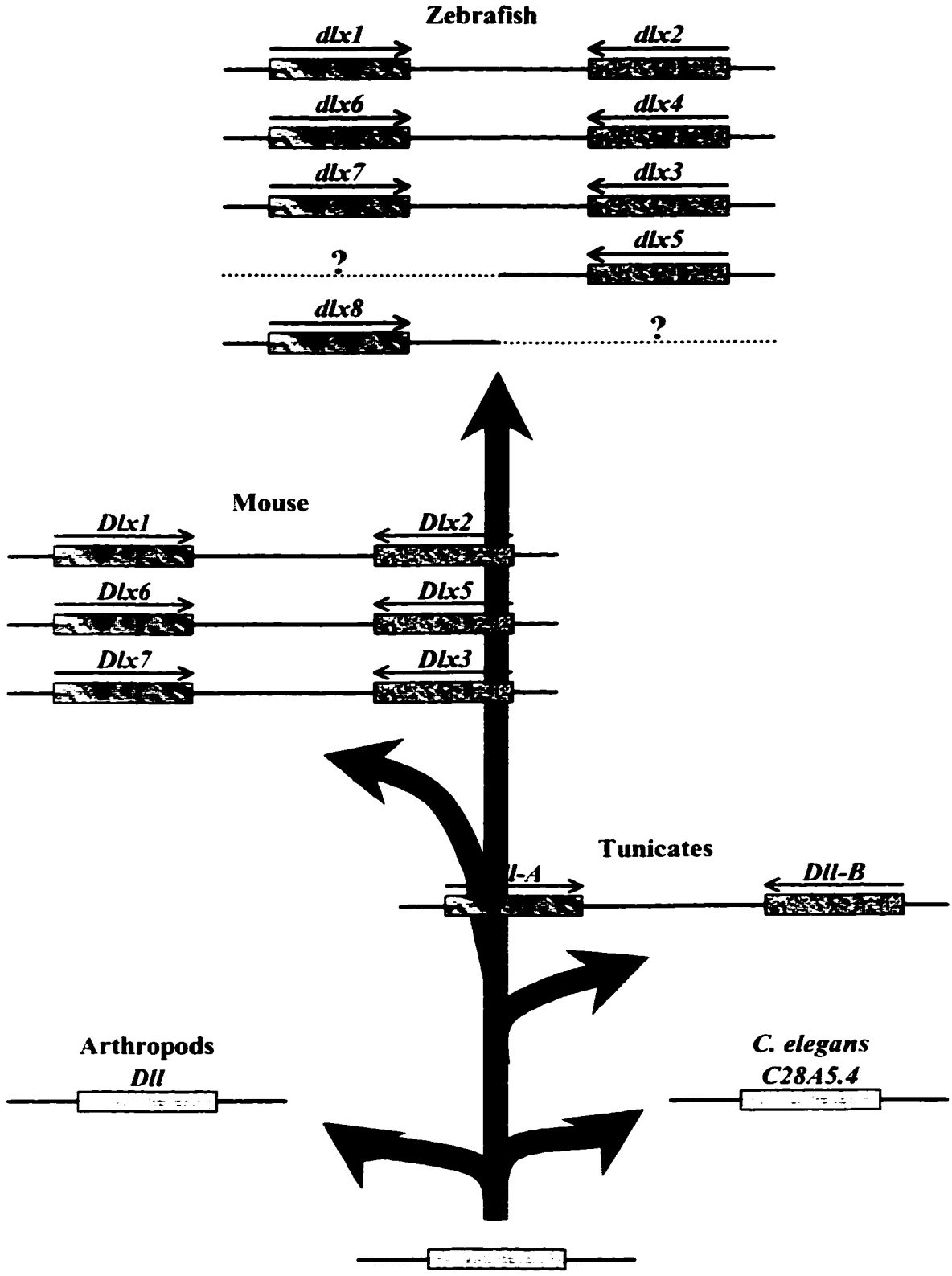
upon these modifications results in a new gene with new functions. Such an occurrence may be observed in the evolution of the vertebrate *Dll*-related gene family.

The vertebrate *Dll*-related family (*Dlx*) consists of 6-8 genes, which are defined by similarities in their homeoboxes to that of *Drosophila Dll*. Members of the *Dlx* family have been identified in all vertebrates examined including human (Nakamura et al., 1996; Scherer et al., 1994; Simeone et al., 1994), mouse (Nakamura et al., 1996; Porteus et al., 1991; Price et al., 1991; Robinson and Mahon, 1994; Robinson et al., 1991; Simeone et al., 1994; Stock et al., 1996; Weiss et al., 1995), rat (Zhao et al., 1994), chick (Ferrari et al., 1995), newt (Beauchemin and Savard, 1992), axolotl (Mullen et al., 1996), the frogs *Xenopus laevis* (Asano et al., 1992; Dirksen et al., 1993; Papalopulu and Kintner, 1993) and *Eleutherodactylus coqui* (Fang and Elinson, 1996), and zebrafish (Akimenko et al., 1994; Ekker et al., 1992; Stock et al., 1996). The *Dlx* genes are organized into 6 paralagous groups based on sequence similarities of the homeodomains which they encode (Table 1.1). The *Dlx* genes can also be grouped into one of two clades based on similarities in sequence motifs within the proteins they encode including the homeodomain as well as regions in the carboxyl terminus (Stock et al., 1996). One clade is made up of the *Dlx1*, *Dlx6* and *Dlx7* paralagous groups and the other, the *Dlx2*, *Dlx3*, and *Dlx5* paralagous groups. The proteins encoded by members of the clade of *Dlx2*, *Dlx3* and *Dlx5* also contain a short highly conserved amino acid sequence in the amino-terminus (Akimenko et al., 1994). In addition, at least 6 of the known *Dlx* genes are organized as inverted convergently transcribed gene pairs, with each gene pair containing one gene from each clade (Fig. 1.3). Thus *Dlx1* and *Dlx2* are a linked pair as are *Dlx5*

Table 1. *Dlx* paralagous groups

Dlx1	*Dlx6*	*Dlx7*
Human *DLX1*	Human *DLX6*	Human *DLX7*
Mouse *Dlx1*	Mouse *Dlx6*	Mouse *Dlx7*
Zebrafish *dlx1*	Zebrafish *dlx6*	Zebrafish *dlx7*
E. coqui EcDlx1	*Xenopus Xdll*	Zebrafish *dlx8*
		Newt *NvHBox5*
Dlx2	*Dlx5*	*Dlx3*
Human *DLX2*	Human *DLX5*	Human *DLX3*
Mouse *Dlx2*	Mouse *Dlx5*	Mouse *Dlx3*
Zebrafish *dlx2*	Zebrafish *dlx4*	Zebrafish *dlx3*
Zebrafish *dlx5*	*Xenopus X-dll3*	*Xenopus X-dll2*
Xenopus X-DLL1	Chicken *Dlx5*	*Xenopus Xdll-2*
Xenopus X-dll4	Rat *rDlx*	Newt *NvHbox4*
E. coqui EcDlx2	*E. coqui EcDlx4*	Axolotl *Dlx-3*
		E. coqui EcDlx3

Figure 1.3. A proposed model for the evolution of the *Distal-less*-related genes as schematized through what is known about *Distal-less* gene number and organization in a number of modern metazoan species. Genes are shown as boxes and likely paralogies are represented by conserved shading within each box. Each gene is labeled and the so far undetermined presence of paired partners for the zebrafish *dlx5* and *dlx8* genes is indicated by question marks. The directions of transcription of the genes are represented by arrows above each gene. The branched arrows indicate branching of metazoan lineages in reflection of the evolution of the *Distal-less*-related genes shown for each species, and are not reflective of the historical timing of these events. The arrows indicate a net increase in complexity based on number of *Distal-less*-related genes and *Dlx* gene pairs and are in no way meant to suggest fitness.



Hypothetical common ancestor to nematode, arthropod and vertebrate *Distal-less* genes

(zebrafish *dlx4*) and *Dlx6* as well as *Dlx3* and *Dlx7* (Ellies et al., 1997b; McGuinness et al., 1996; Nakamura et al., 1996; Ozcelik et al., 1992; Simeone et al., 1994). As is seen for the *Hox* genes, zebrafish has additional *Dlx* genes, *dlx5* and *dlx8*, which are not linked to each other and that do not appear to exist in other vertebrates (Stock et al., 1996).

Evolution Of The *Distal-Less*-Related Gene Family

The relationship between the *Hox* genes and *Dlx* genes is not necessarily relegated to the superficial association of zebrafish simply containing a larger complement of either of these genes as compared, for example, to mammals. Recently it was proposed that the *Dll*-related genes arose during the same duplication events that gave rise to the *Hox* clusters, as a result of linkage to the *Hox* genes (Stock et al., 1996). Thus, the *Dlx1/Dlx2* paralagous pair is linked to the *Hoxd* (zebrafish *hoxda*) cluster, the *Dlx5/Dlx6* pair is linked to the *Hoxa* (zebrafish *hoxaa*) cluster, and the *Dlx3/Dlx7* pair is linked to the *Hoxb* (zebrafish *hoxbb*) cluster (Amores et al., 1998; Nakamura et al., 1996; Rossi et al., 1994; Simeone et al., 1994; Stock et al., 1996). For the 2 additional *Dlx* genes in zebrafish, *dlx8* is linked to the *hoxba* cluster and interestingly *dlx5* maps to the genomic region which is thought to have arisen during the duplication of the *Hoxd* cluster (Amores et al., 1998). Thus *dlx5* seems to have been linked to an ancestral *hoxdb* cluster which has since been lost in zebrafish.

The genomic organization and linkage to the *Hox* genes of the *Dll*-related gene family allows the development of a scenario by which this family likely evolved (Fig. 1.3). An ancestral *Dll*-like gene gave rise to the single *Dlx* homologue observed, for example, in *Drosophila* as already described, or the *C28A5.4* gene in *C. elegans*. The *C28A5.4* gene is distantly linked to the primitive *HOM-C/Hox* cluster and located on the side of the cluster corresponding to the most posteriorly expressed gene (A *C. elegans* Data Base on the world wide web as cited in Stock et al., 1996). The *Dll* gene of arthropods arose from this same ancestor, although it appears that some time after the divergence of the lineage which would give rise to vertebrates, a genomic event in invertebrate evolution resulted in the separation of *Dll* and the *HOM-C* as these genes do not map to the same chromosome in *Drosophila* (Cohen et al., 1989). A gene duplication event resulted in an ancestral convergently transcribed gene pair early in the evolution of chordates. The organization of the *Dlx* homologues in tunicates such as *C. intestinalis* (Di Gregorio et al., 1995) is shown and is representative of a highly primitive chordate, or urochordate. Linkage of this ancestral pair to the single *Hox* cluster in tunicates is unknown, however as there is linkage between the *Dlx* pairs and *Hox* genes in vertebrates, it can likely be assumed that the ancestral *Dlx* pair was linked to the ancestral vertebrate *Hox* cluster. In the lineage which would give rise to vertebrates, a series of duplication events produced at least 3 pairs of *Dlx* genes as shown for the mouse. The *Hoxc* cluster apparently has no *Dlx* gene pair linked to it. It remains unclear if this is due to a loss of *Dlx* genes, or if the *Dlx* pair linked to the ancestral *Hoxcd* cluster, which gave rise to the *Hoxc* and *Hoxd* clusters (Amores et al., 1998) was simply not included in that duplication event. Similarly, the *Evx* gene which is closely linked to the posteriorly

expressed *Hox* genes as well, is also absent from the genomic region of the *Hoxc* cluster in all modern vertebrates examined. The duplication event which produced the additional *Hox* clusters in zebrafish apparently also resulted in at least 2 *Dlx* genes not present in other vertebrates such as mammals. It is known that these 2 genes do not represent a single pair (Stock et al., 1996), but it remains unclear if they are linked to as yet undiscovered *Dlx* genes.

Expression Patterns And Potential Roles Of The Vertebrate *Dlx* Genes

The expression patterns of the zebrafish *dlx* genes are summarized in table 1.2. Based on these patterns of expression, the *Dlx* genes may be inferred to play roles in development of the ventral forebrain (Akimenko et al., 1994; Papalopulu and Kintner, 1993; Porteus et al., 1991; Price et al., 1991; Robinson et al., 1991), olfactory placodes (Akimenko et al., 1994), branchial arches (Akimenko et al., 1994; Dollé et al., 1992; Papalopulu and Kintner, 1993; Robinson and Mahon, 1994; Zhao et al., 1994), otic vesicle and inner ear (Akimenko et al., 1994; Ekker et al., 1992; Papalopulu and Kintner, 1993; Robinson and Mahon, 1994) and limbs/fins (Akimenko et al., 1994; Dollé et al., 1992; Morasso et al., 1995). In general, the patterns of *Dlx* expression during development are relatively well conserved between different vertebrate species. Two striking aspects of the *Dlx* expression patterns are their combinatorial nature and furthermore, their virtual complete overlap between members of each pair. Thus, whatever complement of an organism's *Dlx* genes may be co-expressed in a specific region of the embryo, it appears there is generally an overlap of each member of a pair.

Table 2. Summary of expression of the *dlx* genes during zebrafish development

Ectodermal stripes in gastrula			*dlx3*	*dlx7*				
Ventral forebrain	*dlx1*	*dlx2*			*dlx4*	*dlx6*	*dlx5*	
Olfactory placodes			*dlx3*	*dlx7*	*dlx4*	*dlx6*		
Migrating neural crest		*dlx2*						
Visceral arches	*dlx1*	*dlx2*	*dlx3*	*dlx7*	*dlx4*	*dlx6*		*dlx8*
Dorsal otic vesicle			*dlx3*	*dlx7*	*dlx4*	*dlx6*		
Pectoral fin buds	*dlx1*	*dlx2*	*dlx3*	*dlx7*	*dlx4*	*dlx6*	*dlx5*	*dlx8*
Median fin fold	*dlx1*	*dlx2*	*dlx3*	*dlx7*	*dlx4*	*dlx6*	*dlx5*	*dlx8*

Examples of this include the co-expression of *dlx1* and *dlx2* in the branchial arches and *dlx4* and *dlx6* in the forebrain. In addition there is even similarity in the expression of the single *Distal-less* gene identified in amphioxus (Holland et al., 1996). Similar to the zebrafish *dlx* genes, *AmphiDll* is expressed in the presumptive ectoderm of the gastrula as well as in the cerebral vesicle which has been proposed to be homologous to the vertebrate forebrain

Overlapping patterns of expression for a gene family, as is observed for the *Dlx* genes, presents a problem in determining their roles. One of the most powerful tools for studying roles of genes during development is the technique of selective gene inactivation, or "knocking out the gene". However, it is often observed that when genes are selectively inactivated, the resulting phenotypes prove to either be too subtle to detect or to be less spectacular than expected. A common explanation for this is genetic redundancy. The basis of redundancy in this context is essentially, that despite the nullification of one gene, a related gene expressed in similar regions of the embryo is able to, at least partially, functionally compensate for the loss. This is to be especially expected when the expression patterns of the related genes overlap to the extent observed for the *Dlx* family. In addition, while the identification of domains outside the homeodomain of this family has yet to be reported, sequence similarities do exist and the high degree of sequence conservation of the homeodomain (>90%) suggests that the members of this family recognize a similar DNA-binding site. The phenotypes of *Dlx1* or *Dlx2* null mutants are consistent with a predicted functional redundancy. The phenotype of the *Dlx1* mutant mouse includes an abnormal enteric nervous system and

slight defects in cranial neural crest-derived skeletal components derived from the proximal parts of the first and second branchial arches (Qiu et al., 1997). The *Dlx2* mutant mouse exhibits dysmorphologies in derivatives of the first and second branchial arches as well as abnormal forebrain differentiation (Qiu et al., 1995). Mice lacking both *Dlx1* and *Dlx2* however have a more profound phenotype. This phenotype includes a greater degree of abnormalities of the craniofacial bones (Qiu et al., 1997), a complete absence of maxillary molars (Qiu et al., 1997; Thomas et al., 1997) and striking abnormalities in development of the striatal subventricular zone, in the differentiation of striatal matrix neurons and in the migration of neocortical interneurons from the subcortical telencephalon (Anderson et al., 1997b; Qiu et al., 1997). The importance of *Dlx* genes in craniofacial development is also implied by the observation that craniofacial dysmorphologies in zebrafish embryos treated with exogenous retinoic acid coincide with a complete loss of *dlx* gene expression in the branchial arches (Ellies et al., 1997a). Recently, the murine *Dlx3* gene was implicated in an apparently unique, amongst *Dlx* genes, extra-embryonic role in morphogenesis of the placenta (Morasso et al., 1999) indicating that redundancy can not be assumed in all *Dlx* inactivation experiments.

Studies making use of selective gene inactivations as described above are useful for studying the overall role a gene may play in an organism. The specific role that the gene product plays at the molecular level can rarely be deduced from such experiments. Because of the presence of the homeodomain, Dlx proteins are implicated as regulators of transcription; however based on the available literature, downstream targets of Dlx proteins are largely unknown as is the specific DNA sequence recognized by these

proteins and their requirements for cofactors. One study suggests the ability of Dlx proteins to bind a specific sequence in an enhancer of the *Wnt-1* gene in mouse, which contains two ATTA/TAAT core sequences, and activate transcription (Iler et al., 1995). Unfortunately, the physiological significance of these results remains unclear, as the authors of this study suggest this enhancer represents a negatively regulated element responsible for repressing *Wnt-1* expression in the developing forebrain. Another study suggests that in rat, Dlx5 is able to repress expression of the *osteocalcin* gene (Ryoo et al., 1997). This implicates a role for a Dlx protein in osteoblast differentiation and bone morphogenesis, which may explain the relationship between craniofacial dysmorphologies of the skeleton and loss of *Dlx* expression (Ellies et al., 1997a; Qiu et al., 1997; Qiu et al., 1995).

The molecular mechanisms by which Dlx proteins exert their influences on transcription are also largely unknown. It has been suggested however that specific Dlx proteins are able to form homodimers as well as heterodimers with other Dlx proteins and members of the Msx family of homeodomain proteins (Zhang et al., 1997). These dimerizations are mediated by the homeodomains of both these families and are independent of DNA-binding. The authors propose that Msx proteins are transcriptional repressors that function without requiring to bind DNA, whereas Dlx proteins are DNA-binding-dependant transcriptional activators. The mechanism by which Msx proteins may repress transcription independent of DNA binding is not addressed. However, they suggest that the repressor function of Msx proteins can be abolished by their binding to Dlx proteins. This indicates that, in addition to being able to activate transcription by

binding *cis*-regulatory elements, Dlx proteins can also antagonize the repressor function of Msx proteins by interacting with them independent of DNA-binding. The physiological significance of Dlx proteins forming dimers with other Dlx proteins is not made clear.

The factors that regulate the vertebrate *Dlx* genes are largely unknown as well, as are any elements which may direct expression of the *Dlx* genes during development. One exception is an approximately 1 kb region from directly upstream of the *Xenopus Xdll-2* gene which has been determined to direct expression in the epidermis of frog embryos and which can also specify expression in murine epidermal cells (Morasso et al., 1994).

Statement Of Problem

The purpose of the project described in this thesis was to study the regulation of the *Dlx* family during development. As stated above, very little is known about the genetic hierarchies, both upstream and downstream, in which *Dlx* genes are involved. A number of explanations can be forwarded for the striking patterns of overlapping expression observed for the *Dlx* genes. One possibility is that Dlx proteins, as part of their function may be involved in regulating *Dlx* gene expression. Precedents exist for this type of cross-regulation such as that already discussed for the *Hox* genes. Another possibility, applicable to the paired *Dlx* genes which share virtually identical expression patterns, is that the paired *Dlx* genes share *cis*-acting regulatory elements. Precedents exist for this as well such as enhancer sharing which has also been already introduced

using the example of the *Hox* genes. A logical place to search for such shared enhancers is the intergenic regions separating the *Dlx* gene pairs. A number of such elements have been identified and I have demonstrated the ability of Dlx proteins to bind and activate transcription through a subset of these elements. I have also demonstrated that sites of auto- and/or cross-regulation are not restricted to intergenic enhancers as Dlx3 has the ability to activate transcription through a region of DNA from directly upstream of *dlx4*. In addition, by studying the activity of the intergenic enhancers in metazoans as diverse as mammals and insects, the picture of the evolution of this gene family has not only become clearer, but a better understanding of how gene families made up of multiple members may have evolved has also been achieved.

2. CROSS-INTERACTIONS BETWEEN TWO MEMBERS OF THE *DLX* FAMILY OF HOMEOBOX-CONTAINING GENES DURING ZEBRAFISH DEVELOPMENT

T. Zerucha*, J.-P. Muller*, N. Chartrand and M. Ekker[1]

T. Zerucha and N. Chartrand. Ottawa Civic Hospital Loeb Research Institute, Ottawa, ON, Canada, K1Y 4E9. Department of Cellular and Molecular Medicine, University of Ottawa.

J.-P. Muller. Laboratoire d'Embryologie Moléculaire et Expérimentale, URA 2227 du CNRS, Université de Paris-Sud XI, 91405 Orsay, France.

M. Ekker. Ottawa Civic Hospital Loeb Research Institute, Ottawa, ON, Canada, K1Y 4E9. Department of Medicine and Department of Cellular and Molecular Medicine, University of Ottawa.

* The two authors contributed equally to this work.

[1]Author to whom all correspondence should be addressed. (Voice: (613) 798-5555 ext 6033; Facsimile: (613) 761-5036; e-mail: ekker@mail.civich.ottawa.on.ca).

Published: Biochemistry and Cell Biology 75: 613-622 (1997)

ABSTRACT

The *Dlx* homeobox genes of vertebrates are transcribed in multiple cells of the embryo with overlapping patterns but often with different onsets of expression. Here, we describe the interaction between two *dlx* genes, *dlx3* and *dlx4*, during zebrafish development. The observation that *dlx3* expression precedes that of *dlx4* in the otic vesicle led us to investigate whether *dlx3* had the ability to control expression of *dlx4*. Truncated versions of *dlx3* were over-expressed in zebrafish embryos and the expression patterns of *dlx4* were examined later in development. Over-expression of truncated forms of Dlx3 or of a Dlx3-Dlx2 chimera were found to result in perturbations in *dlx4* expression. In addition, co-transfection experiments, in cultured cells, indicated the ability of Dlx3 to activate transcription through the 5' flanking region of *dlx4*. These results suggest that *dlx4* is one of the target genes of *dlx3* in embryos and that cross-regulatory interactions between *Dlx* genes may be one of the mechanisms responsible for their overlapping expression.

Key Words: Danio rerio, *branchial arches, inner ear, homeodomain, transgenic animals.*

RÉSUMÉ

Les gènes à boîte homéo de la famille *Dlx* des vertébrés sont co-exprimés dans de nombreuses cellules de l'embryon. Nous rapportons l'interaction entre deux gènes *dlx*, *dlx3* et *dlx4* du danio (poisson-zèbre). L'expression de *dlx3* précède celle de *dlx4* dans les vésicules otiques au cours de leur développement, ce qui nous a amené à déterminer si *dlx4* pouvait constituer un des gènes cibles de *dlx3*. Nous avons sur-exprimé des ARN messagers codant pour des versions tronquées de Dlx3 dans des embryons de danio et avons mesuré l'expression de *dlx4* à des stades ultérieurs de développement. La sur-expression des versions tronquées de Dlx3 ou d'une chimère Dlx3-Dlx2 causèrent des perturbations de l'expression de *dlx4*. De plus, des expériences de co-transfection ont démontré la capacité de Dlx3 d'activer la transcription d'un gène "reporter" contenant 1.7kb de séquence 5'-flanquante de *dlx4*. Ces résultats suggèrent que *dlx3* est un des activateurs de *dlx4* chez l'embryon et que des interactions croisées entre les gènes *Dlx* sont, du moins en partie, responsables de leur co-expression au cours du développement.

Mots-Clés: Danio rerio, *arcs branchiaux, oreille interne homeodomaine, animaux transgéniques.*

INTRODUCTION

Transcription factors encoded by homeobox genes are found throughout the eukaryotic kingdoms and play important roles during development. In metazoans, these genes can be divided into two general classes based on similarities in sequence and organization in the genome. The *Hox* genes are organized in clusters in the genome and are involved in specifying the antero-posterior patterning of the organism. The cluster organization of the *Hox* genes was shown to influence their function. Members of the other class of homeobox-containing genes are spread throughout the genome. They are divided into families on the basis of the sequence of their homeobox. Members of the *Dlx* gene family in vertebrates are defined by having homeoboxes similar in sequence to that of the *Drosophila Distal-less* (*dll*) gene (Cohen et al., 1989; Vachon et al., 1992). *Dlx* genes have been identified in rat (Zhao et al., 1994), *Xenopus* (Asano et al., 1992; Dirksen et al., 1993; Papalopulu and Kintner, 1993), the frog *Eleutherodactylus coqui* (Fang and Elinson, 1996), newt (Beauchemin and Savard, 1992), axolotl (Mullen et al., 1996, and C. Nolte, J. Armstrong and M. Ekker, unpubl. observations), chicken (Ferrari et al., 1995), mouse (Nakamura et al., 1996; Porteus et al., 1991; Price et al., 1991; Robinson and Mahon, 1994; Robinson et al., 1991; Simeone et al., 1994; Weiss et al., 1995), zebrafish (Akimenko et al., 1994; Ekker et al., 1992; Stock et al., 1996), and human (Nakamura et al., 1996; Scherer et al., 1995; Simeone et al., 1994). At present, as many as six different members of this family have been identified in most vertebrate species with the exception of zebrafish in which eight different members have been identified (Akimenko et al., 1994; Ekker et al., 1992; Stock et al., 1996).

Little is currently known, however, of the role that the *Dlx* genes play. During zebrafish development, *dlx* genes are expressed in: ectodermal stripes in the gastrula; the ventral forebrain; the olfactory placodes; migrating neural crest; the visceral arches; the otic vesicle; the pectoral fin buds; and the median fin fold (Akimenko et al., 1994; Ekker et al., 1992; Ellies et al., 1997a,b). Although changes in *dlx* expression have been observed in some of the mutants of the large screen carried out in zebrafish (see for example (Malicki et al., 1996)), none of the mutants was identified as carrying a mutation in a *dlx* gene specifically. Selective inactivation of *Dlx2* (Qiu et al., 1995), *Dlx1* or both *Dlx1* and *Dlx2* (Qiu et al., 1997) in the mouse results in dysmorphologies of cranio-facial cartilage elements. Similarly, the loss of *dlx* expression is observed to coincide with cranio-facial dysmorphologies in zebrafish embryos treated with retinoic acid (Ellies et al., 1997a). All of these studies suggest that *Dlx* genes are involved in the development of skeletal elements of the head. However, several regions of the embryo where *Dlx1* and/or *Dlx2* are expressed in mouse embryos were largely unaffected in the null mutants. One of the explanations for the lack of mutant phenotype in these regions of the embryo where *Dlx* genes are expressed is functional redundancy. Vertebrate *Dlx* genes are generally expressed with overlapping patterns. Thus it is thought that loss of expression of one *Dlx* gene in a specific region of the embryo can be compensated, at least partially, by the other *Dlx* gene(s) also expressed in that region. In addition to overlapping patterns of expression, the *Dlx* genes are also expressed with different temporal onsets in specific regions of the embryo. For example, in the visceral arches of zebrafish embryos, all members are expressed with the exception of *dlx5* (Ellies et al., 1997b), however, *dlx2* is

expressed first, followed by the other *dlx* genes (Akimenko et al., 1994; Ellies et al., 1997b).

To understand the molecular mechanisms responsible for the overlapping patterns of *dlx* expression in zebrafish embryos, we are examining the possibility that zebrafish *dlx* genes may cross-regulate each other. This hypothesis is based on the temporal expression patterns of the *dlx* genes, and the fact that this type of regulation is observed with other homeobox-containing gene families (Arcioni et al., 1992; Faiella et al., 1994; Zappavigna et al., 1991). For this study we have chosen to examine the interaction between the *dlx3* and *dlx4* genes. During development of the inner ear in zebrafish, *dlx3* is one of the first genes to be expressed in the presumptive otic placode along with *dlx7* (Akimenko et al., 1994; Ekker et al., 1992). This expression of *dlx3* is followed by *dlx4* in the same cells of the otic vesicle (Akimenko et al., 1994). To examine the role of the Dlx3 protein in *dlx4* expression, full length and truncated versions of *dlx3* were over-expressed in zebrafish embryos by injecting *in-vitro* transcribed mRNA coding for these transcripts into single cell embryos, and the expression of *dlx4* examined. Truncated versions of Dlx3, both containing and lacking the homeodomain were found to perturb *dlx4* expression as was a chimeric construct consisting of the amino-terminal half of Dlx3 and the carboxy-terminal half of Dlx2. In addition, co-transfection experiments in cultured cells showed that Dlx3 can activate transcription through a 1.7 kb genomic DNA fragment from the upstream region of *dlx4*. Taken together these results suggest that *dlx3* can act as a positive regulator of *dlx4* during development and that cross-regulatory

interactions between *dlx* genes contribute to their overlapping patterns of expression in embryos.

MATERIALS AND METHODS

Animals

Adult zebrafish were obtained from a local supplier and were maintained according to standard procedures (Westerfield, 1995). Embryos were staged at 28.5° C according to hours (h) and days (d) post fertilization.

In Vitro Transcription Of Synthetic mRNA

Synthetic mRNA coding for full length or truncated versions of Dlx3 were prepared using convenient restriction sites in the coding region of the *dlx3* gene. Truncated versions of *dlx3* as well as a *dlx3-dlx2* chimera were prepared and are shown schematically in Figure 2.1. RNA*i* codes for full length Dlx3 protein. RNA*ii* was created using a *Hind*III site at the 3' end of the *dlx3* homeobox and codes for a 182 amino acid polypeptide consisting of the amino-terminus and homeodomain. RNA*iii* was created using a *Xho*I site in the 5' region of the *dlx3* homeobox and codes for a 131 amino-acid polypeptide including the amino-terminus and first 9 amino acids of the homeodomain. RNA*iv* was created using a *Pst*I site in the 5' region of the *dlx3* gene and codes for the 99 amino-terminal amino acids of Dlx3. A Dlx3-Dlx2 chimera, encoded by

RNA*v*, was also created by making use of the conserved *Xho*I site in the homeoboxes of *dlx2* and *dlx3*. The creation of this chimera does not affect the reading frame of *dlx2* nor *dlx3*. The chimeric protein consists of the amino-terminus and first 9 amino-acids of the homeodomain of Dlx3, the remainder of the homeodomain and carboxy-terminus are coded for by *dlx2*.

The plasmid vector for all templates is the pBluescript plasmid (Stratagene). Templates were linearized with the restriction enzymes that coincide with the 3'-end of the *dlx* cDNA sequence in each clone. *In vitro* transcription and capping of RNA were carried out from 1 μg of linearized template using the "mCAP" mRNA capping kit (Stratagene) according to the manufacturer's instructions. Synthetic RNA was quantified by absorbance at 260 nm.

Injection Of mRNA Into Developing Zebrafish

Synthetic, capped RNA was injected into 1 cell stage zebrafish embryos essentially as described in Stuart et al., 1988. Briefly, synthetic capped RNA was dissolved in distilled water containing 0.1% phenol red to visualize and estimate the volume of RNA solution injected. Embryos were then kept for a period of 27 h in embryo medium (Westerfield, 1995). A control experiment in which a full-length *dlx3* synthetic mRNA labeled with digoxygenin was injected showed that this RNA was uniformly distributed after 24 h.

In Situ Hybridization

Embryos were fixed at 27 h in a phosphate buffered saline solution (PBS) containing 4% paraformaldehyde. *In situ* hybridizations were performed on whole mount embryos using antisense riboprobes as previously described (Akimenko et al., 1994). The anti-sense riboprobes for *dlx2*, *dlx3*, and *dlx4* were transcribed from the previously reported, 1667 bp, 1532 bp and 1123 bp cDNAs, respectively (Akimenko et al., 1994).

Plasmids And Transient Co-Transfection Experiments

To construct the reporter plasmid for co-transfection assays, a 1.7 kb *Xba*I-*Pst*I fragment of the *dlx4* 5'-flanking region (Ellies et al., 1997b) was cloned upstream of the minimal thymidine kinase (tk) promoter driving the chloramphenicol acetyltransferase (CAT) gene in the pBLCAT2 plasmid (Luckow and Schütz, 1987). The position of the *Pst*I site in the *dlx4* genomic fragment is about 100 bp upstream of the 5'-end of the *dlx4* cDNA (Akimenko et al., 1994). The full length *dlx3* cDNA was cloned in the pTL2 expression vector under the control of the SV40 early promoter. Both the reporter and expression vectors contain the SV40 polyadenylation signal.

Transient co-transfection experiments were performed in the P19 murine cultured embryonic carcinoma cell line. Cells were plated 24 hours prior to transfection at a density of $5x10^5$ cells per 100 mm dish. Transfections were carried out by calcium phosphate precipitation (Sambrook et al., 1989). A total of 10 μg of plasmid DNA was

used in each transfection. This included: 2 µg pRSV- ßgal as an internal control for transfection efficiency; 2 µg of the reporter plasmid, 2 µg of effector plasmid, and sheared calf thymus DNA to 10 µg. Precipitates were left on the cells for 16 h, and the cells were harvested 60 h post-transfection. Cells were harvested in PBS, pelleted by centrifugation and resuspended in freeze/thaw buffer (250 mM Tris-HCl pH 8, 10 mM DTT, 15% glycerol). Whole cell extracts were prepared by repeated freezing and thawing. Beta-galactosidase activity was assayed as described by Sambrook et al. (Sambrook et al., 1989). CAT activity was determined by thin-layer chromatography and measured as percent conversion of mono- and diacetylated chloramphenicol relative to unmodified plus acetylated chloramphenicol using the BIO-RAD GS-525 Molecular Imager System. CAT activity was standardized to ß-galactosidase levels to compensate for variation in transfection efficiencies. Transfections were performed in duplicate and were repeated a minimum of three times.

RESULTS

Expression Of *dlx4* Is Perturbed In Embryos In Which Truncated Versions Of *dlx3* Are Over-Expressed.

In a first series of experiments, a synthetic *dlx3* RNA coding for a truncated form of the Dlx3 protein which ends at the carboxy-terminal part of the homeodomain (RNA*ii*, Fig. 2.1) was injected in one-cell stage embryos. *Dlx* gene expression was determined 27 h later. In 27 embryos injected with 4 ng of RNA*ii*, 15% exhibited a loss of *dlx4*

expression in the ear (Fig. 2.2E, F), and 17% lost *dlx4* expression in the visceral arches (Fig. 2.2C, D). No perturbation of *dlx4* expression was observed in the forebrain nor median fin fold (Table 2.1). One possible explanation of this effect is that the incomplete Dlx3 protein can compete with the endogenous Dlx3 protein for its binding sites on the DNA, yet is unable to exert its normal function. To determine if the presence of the homeodomain in the truncated polypeptide is important for its effects on *dlx4* expression, a synthetic RNA coding for a truncated form of Dlx3 which lacks the homeodomain was injected into single cell zebrafish embryos. In 105 embryos injected with 1 ng of RNA*iii*, 19% exhibited a loss of *dlx4* expression in the ear (Fig. 2.3C; Table 2.1), 2.5% of the embryos lost *dlx4* expression in the visceral arches (Fig. 2.3B) and 1% lost expression in the forebrain and arches (Table 2.1). As in all other experiments described in this report, *dlx4* expression in the median fin fold was unaffected in embryos injected with RNA*iii* (Fig. 2.3D, E). To determine if the quantity of mRNA injected plays a role in the degree of perturbation of *dlx4* expression, the previous experiment was repeated with 5 ng of RNA. The expression of *dlx4* was examined in 180 embryos, and a phenotype very similar to that produced by the injection of 1 ng of RNA*iii* was observed. The maximal effect was seen in the ear where 10% of the embryos examined had lost expression of *dlx4*. In 3.3% of the remaining embryos examined *dlx4* expression was lost in the visceral arches; and in 2.7%, it was lost in the forebrain, visceral arches and otic vesicles.

The effect of RNA*iii* appears to specifically affect *dlx4* expression, as embryos injected with this RNA showed no changes in the expression of *dlx2* or of the endogenous *dlx3* gene (not shown). Furthermore, we did not observe any significant

delays in development in the embryos injected with RNA*iii*, nor with any other synthetic RNA used in this study, that could account for the changes in *dlx4* expression. Finally, control injections of a synthetic *lac-z* RNA, coding for the enzyme beta-galactosidase, had no effect on *dlx4* expression.

The greatest perturbation in *dlx4* expression was observed in embryos injected with an even shorter *dlx3* synthetic RNA. In 70 embryos injected with RNA*iv*, 37% showed a loss of *dlx4* expression in the ear (not shown), 8.5% also exhibited a loss in *dlx4* expression in the visceral arches and 18.5% lost expression in the forebrain in addition to the visceral arches (not shown).

A Dlx3-Dlx2 Chimeric Polypeptide Also Perturbs *dlx4* Expression During Development

To determine if the carboxy-terminal half of Dlx2 is able to functionally substitute for the same domain of *dlx3*, RNA*v*, which codes for a chimera consisting of the amino-terminal half of Dlx3 and the carboxy-terminal half of Dlx2, was injected into single cell stage zebrafish embryos. This chimeric polypeptide was created by making use of a conserved *Xho*I restriction endonuclease cleavage site in the homeoboxes of *dlx3* and *dlx2*. The creation of this chimera does not disturb the reading frame of *dlx2* and encompasses the amino-terminal half and first 9 amino acids of the homeodomain of Dlx3 and the remaining 51 amino acids of the homeodomain correspond to that of Dlx2 as does the carboxy-terminal half of the polypeptide. The over-expression of the Dlx3-

Dlx2 chimera also perturbs expression of *dlx4*. In 12.5% of the 41 embryos examined, *dlx4* expression is lost in the visceral arches (Fig. 2.4B) and in another 12.5%, expression is lost in the forebrain, visceral arches and ear (Fig. 2.4 C).

Dlx3 Is Able To Regulate Transcription Through The 1.7 Kb 5' Flanking Region Of *dlx4*

To determine if Dlx3 can stimulate *dlx4* expression by interacting with regulatory sequences in the *dlx4* gene, we examined the effect of Dlx3 on a 1.7 kb DNA fragment of the immediate *dlx4* 5'-flanking region in co-transfection experiments using cultured P19 murine embryonic carcinoma cells. Constructs used for transient co-transfection experiments are shown in Fig. 2.5A. The co-transfection of a *dlx3* expression construct with the *dlx4* reporter construct resulted in a 5 fold increase in the activity of the reporter gene relative to its expression in the absence of Dlx3 (Fig. 2.5B).

DISCUSSION

Dlx3 Is A Potential Activator Of *dlx4* Expression During Development

During development of the ear, *dlx3* and *dlx7* are amongst the first genes expressed in the presumptive otic placode, a thickening of the ectoderm which will eventually produce the inner ear (Ekker et al., 1992; Ellies et al., 1997b). Expression of *dlx7* in the otic placode and vesicle however appears to be weaker than that of *dlx3*. The

expression of *dlx4* follows by a few hours that of *dlx3* and *dlx7* in the same cells of the otic vesicle. To determine the role of *dlx3* on the expression of *dlx4*, synthetic mRNA's coding for truncated versions of the Dlx3 protein were over-expressed in developing zebrafish embryos. This type of experiment has effectively been used previously to ubiquitously express protein products in zebrafish embryos (deVries et al., 1996; Kelly et al., 1995a,b). Over-expression of full length *dlx3* had no effect on the expression of *dlx4*. Over-expression of truncations of *dlx3* however, affected expression of *dlx4* to varying degrees in the visceral arches, otic vesicle and ventral forebrain. The expression of *dlx4* was unaffected in all experiments in the median fin fold, regardless of the RNA injected. Furthermore, expression of *dlx2* was unaffected in embryos injected with RNA*iii* (not shown). Generally, the greatest effects on *dlx4* expression were observed in the otic vesicle and in the visceral arches. In the otic vesicle, expression of *dlx3* (and of *dlx7*) precedes that of *dlx4*. The loss of expression of *dlx4* in embryos in which truncated versions of *dlx3* are over-expressed also suggests that one of the target genes for *dlx3* in the early otic vesicle are the *dlx* genes subsequently expressed in the same cells of the vesicle, such as *dlx4* and *dlx6*. This is supported by our observation that *dlx3* is able to activate transcription through a DNA fragment from the immediate *dlx4* 5'-flanking region. Our results also suggest that in the ear, *dlx3* would have a preponderant function over *dlx7* in the regulation of *dlx4* because, in embryos injected with truncated versions of *dlx3*, *dlx7* expression was not affected (Ellies, Muller and Ekker, unpublished observations), whereas *dlx4* expression was impaired (this report).

Injection of full length *dlx3* mRNA into embryos results in ubiquitous expression of this gene but did not produce any ectopic *dlx4* expression. This contrasts with our transfection studies which showed that the Dlx3 protein can up-regulate transcription from a fragment of the *dlx4* 5'-flanking region. The inability of Dlx3 to induce by itself ectopic *dlx4* expression in injected embryos could be attributed to the absence of necessary co-factors or to a closed transcriptional state of the chromatin at the *dlx4* locus. The latter phenomenon would not be observed in a transient co-transfection experiment. It is also possible, although unlikely, that the co-factors necessary for activation of the *dlx4* gene are present in P19 cells. Therefore, our result that Dlx3 is able to stimulate transcription from cis-acting elements found upstream of *dlx4* in P19 cells suggests that Dlx3 is one of the factors that are involved in *dlx4* regulation whereas mRNA injection experiments suggest that Dlx3 is necessary but not sufficient for *dlx4* expression.

Another potential target of the Dlx3 protein in embryos could be the *dlx3* gene itself as many developmental genes have been shown to participate in auto-regulatory positive feedback loops (Pöpperl and Featherstone, 1992). Our observations, however, that injected synthetic RNA*iii* coding for truncated version of Dlx3 had no significant effect on endogenous *dlx3* expression (not shown) would be inconsistent with such a positive feedback loop.

The functional significance of cross-regulatory interactions between *Dlx* genes is not yet clear. The overlapping expression of multiple *Dlx* genes has been suggested as a cause for the absence of, or the mild phenotypes observed in mice homozygous for

targeted null mutations in one or two *Dlx* genes (Qiu et al., 1997; Qiu et al., 1995). The mechanisms that produce specific combinations of *Dlx* genes in distinct cell populations, combinations that are well conserved in evolution (Ellies et al., 1997b), are also unclear at this time. For example, the presence of *cis*-acting Dlx binding sites, as part of an auto-regulatory feedback loop, may result, indirectly, in cross-regulatory mechanisms. This would happen if a Dlx protein, other than the product of the *dlx* gene which is auto-regulated, is able to bind to the auto-regulatory *cis*-acting sequences. Our experiments, however, could not provide evidence for auto-regulation in the case of *dlx3*, as mentioned above.

A Conserved Region Of Dlx3 Is Important For Regulation Of *dlx4*

From the over-expression studies described here, we can suggest regions of Dlx3 that are important for its function. The ability of the amino-terminal region upstream of the homeodomain to interfere with normal Dlx3 function suggests that this region contains a functionally relevant domain, perhaps the 19 aa conserved motif (designated A in Fig. 2.1) specific to the sub-family of Dlx proteins consisting of *Dlx2*, *Dlx3* and *Dlx5* genes (the latter is called *dlx4* in zebrafish) (Akimenko et al., 1994; Stock et al., 1996). The ability of this interference to occur in the absence of the DNA-binding homeodomain suggests a squelching mechanism. Thus, the amino-terminal region would contain a domain involved in interactions with another, yet unidentified factor, and the truncated protein is competing for this factor with the endogenous Dlx3 protein (Fig. 2.6A, B). The truncated polypeptide lacks the ability to recognize a DNA-binding site, and is thus

sequestering this factor away from the Dlx3 target sites in the genome. The ability of a truncated polypeptide which includes the homeodomain (encoded by RNA*ii*) to also interfere with normal Dlx3 function suggests that the carboxyl-terminal region of Dlx3 also contains a functionally important domain, possibly responsible for transcriptional activation or other protein-protein interactions necessary for optimal DNA binding to the target sites (Fig. 2.6C).

In some of the embryos injected with the RNA*v* coding for the Dlx3-Dlx2 chimera (Fig. 2.5) and in some of the embryos injected with RNA coding for the truncated forms of Dlx3, expression of *dlx4* was decreased in the ventral forebrain . It is not possible to explain this information in terms of interference with endogenous *dlx3* function because *dlx3* is not expressed in forebrain cells (Akimenko et al., 1994). It is, however, possible that in those cells, the mutant form of Dlx3 interferes with the function of another *dlx* gene, possibly *dlx2*. Similarly this would suggest that the decreases in *dlx4* expression that we observed in the visceral arches may be due to the interference of the mutant *dlx3* RNA with both Dlx3 and Dlx2 function. Both the Dlx3 and Dlx2 proteins share the 19 aa conserved motif near the amino-terminus. We suggest that the putative factor(s) (Fig. 2.6) that interact with this motif may be necessary for Dlx function in several contexts and would interact with one or several Dlx proteins depending on the cell type. Consistent with this model, over-expression of synthetic RNA coding for a truncated version of Dlx2 caused decreases in *dlx4* expression, predominantly in the visceral arches (Muller and Ekker, unpublished observations).

Dlx Genes Are Not Created Equal

Despite the similarities between Dlx proteins, specifically Dlx2 and Dlx3, in the homeodomain (56 identical aa out of 61) as well as in the 19 amino acid region A (Fig. 2.1), we have observed that a chimera consisting of the amino-terminal half of Dlx3 and carboxyl-terminal half of Dlx2 interferes with the regulation of *dlx4* expression (Fig. 2.4) in a manner similar to truncated Dlx3 while over-expression of the entire Dlx3 has no effect. Mechanisms can be proposed to explain this observation. First, the Dlx3-Dlx2 chimera competes with Dlx3 for binding sites on the DNA but cannot make the proper protein-protein contacts necessary for *dlx4* activation. This would imply that Dlx3 contains a specific domain downstream of its homeodomain that is not only functionally significant, as suggested in the preceding section, but is also specific in that it is not shared by Dlx2. This model assumes that the Dlx3-Dlx2 chimera has the same DNA binding specificity as Dlx3. The homeodomain of the chimera is more similar to that of Dlx2 than to that of Dlx3 because 4 of the 5 differences between the Dlx3 and Dlx2 homeodomains are found downstream of the *Xho*I site used to make the chimeric cDNA (Fig. 2.1). It is not known at present if any of these 4 amino acids influences the DNA binding specificities of the two proteins. Furthermore, amino acids downstream of the Dlx2 homeodomain may influence DNA binding specificity and target the Dlx3-Dlx2 chimera to sites in the genome different from those preferred by Dlx3. Therefore, it is still possible that the chimera cannot bind to the Dlx3 target sites, including, presumably, sequences in the *dlx4* 5'-flanking region fragment that is activated by Dlx3 in cell culture.

In that case, it would act as the truncated forms of Dlx3 and sequester a factor that interacts with the amino terminal region of Dlx3.

Our studies provide a possible explanation as to how members of the *Dlx* family may be able to exert different effects particularly during early development of specific tissues such as the otic vesicle and visceral arches. While Dlx proteins share certain functional specificities as evidenced by their apparent functional redundancy observed in gene inactivation studies (Qiu et al., 1997; Qiu et al., 1995), they also contain specific regions which confer upon them some level of uniqueness. In addition this is the first work suggesting the functional significance of regions of a Dlx protein outside of the homeodomain. Based on amino acid sequence analysis, Dlx protein domains of potential functional significance include the 19 aa conserved element mentioned above as well as a tyrosine-rich region upstream of the homeodomain and two serine-rich regions downstream of the homeodomain (Akimenko et al., 1994; Shirasawa et al., 1994). From the work described here it would also seem that in addition to these regions which are conserved amongst Dlx proteins other protein domains that remain to be characterized confer each Dlx protein with unique functional characteristics.

ACKNOWLEDGEMENTS

We thank Lynda Laforest for expert technical assistance, and Marie-Andrée Akimenko for advice and critical reading of the manuscript. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to M. E. and a province of Quebec FCAR studentship to N. C.

Figure 2.1 Summary of synthetic mRNA coding for full length, truncated and chimeric Dlx3 used for injections in zebrafish embryos. The restriction sites used to create the truncations and chimeric Dlx3-Dlx2 protein are shown. RNA*i* codes for full length Dlx3. RNA*ii* codes for a polypeptide of 182 aa in length consisting of the amino-terminal region of Dlx3 as well as the homeodomain (HD). RNA*iii* codes for a polypeptide 131 aa in length consisting of the amino-terminal region of Dlx3 as well as the first 9 amino acids of the homeodomain. RNA*iv* codes for a polypeptide 99 aa in length consisting of the amino-terminal region of Dlx3 (24 aa upstream of the homeodomain). RNA*v* codes for a chimeric polypeptide consisting of the amino-terminal half and first 9 aa of the homeodomain of Dlx3 and the carboxyl-terminal half and majority of the Dlx2 homeodomain. The highly conserved 19 aa region present near the amino-terminus in the sub-family of Dlx proteins consisting of Dlx2, Dlx3, and Dlx5 (Dlx4 in zebrafish), is labelled A.

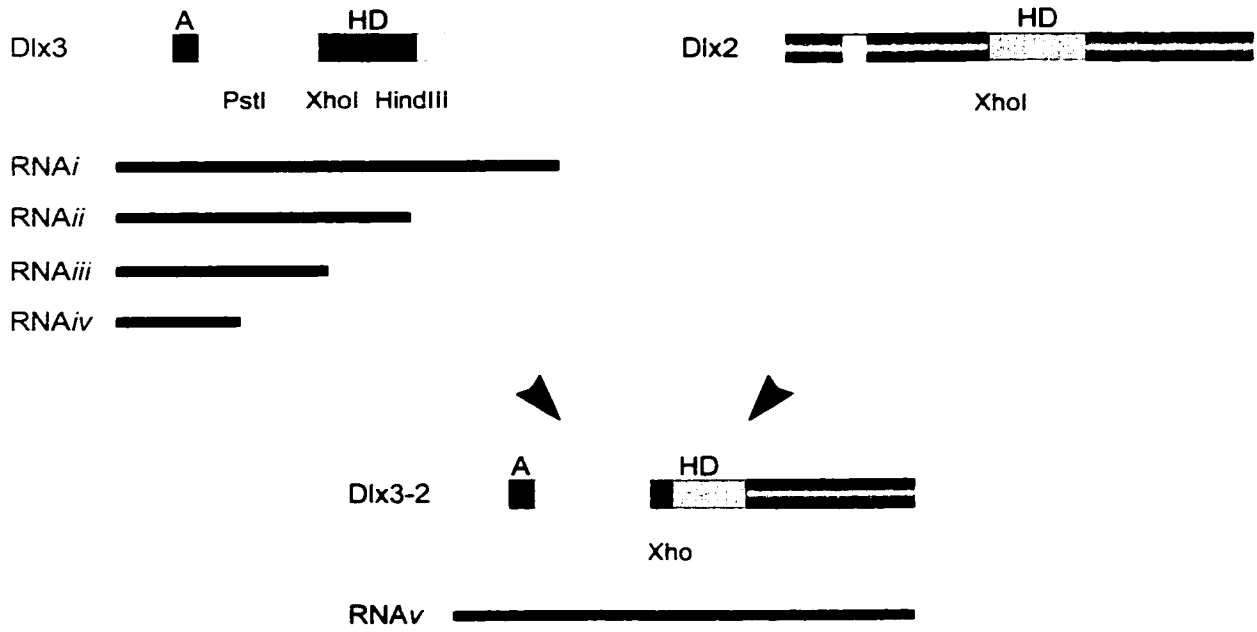

Dlx3
A
HD
PstI
XhoI
HindIII
Dlx2
HD
XhoI
RNAi
RNAii
RNAiii
RNAiv
Dlx3-2
A
HD
Xho
RNAv

Figure 2.2 Perturbations in expression of *dlx4* as a result of the over-expression of the truncation of *dlx3* encoded by RNA*ii*:. **A**, **C**, **E**, Sagittal; **B**, **D**, **F**, dorsal views of *dlx4* expression in the forebrain (f), visceral arches (va) and otic vesicle (o) of control embryos. **A-B**, wild-type *dlx4* expression. **C-D**, 17% of embryos injected with RNA*ii* show perturbed expression of *dlx4* in the visceral arches as indicated by the white arrowheads. **E-F**, 15% of the embryos injected with RNA*ii* show loss of expresion of *dlx4* in the otic vesicle as indicated by the arrowhead(s). Scale bar: 190 μm.

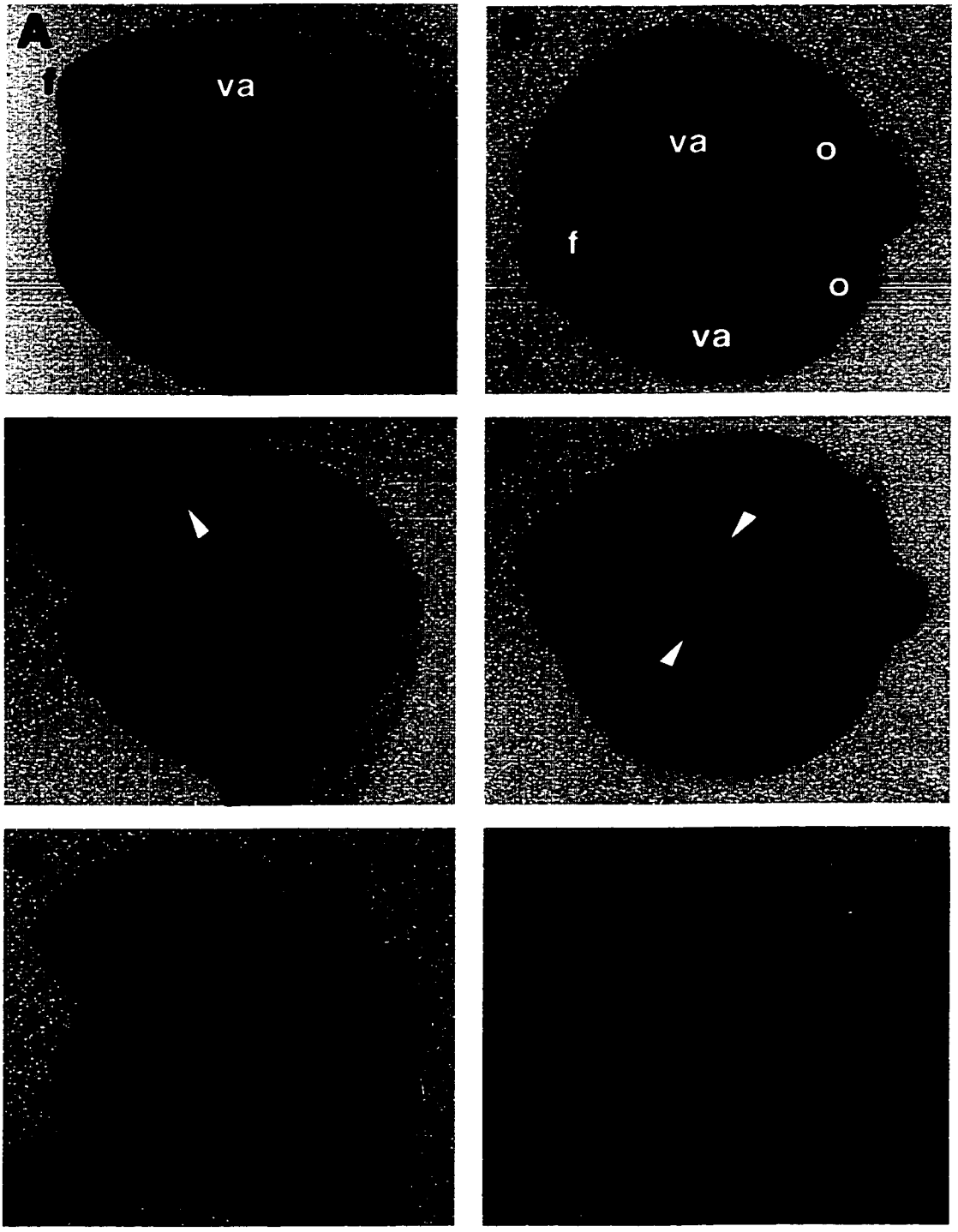

A
va
va
o
f
o
va

Figure 2.3 Perturbations in *dlx4* expression resulting from over-expression of the Dlx3 truncation encoded by RNA*iii*. **A**, wild-type *dlx4* expression. **B**, 2.5% of embryos in which the truncation of Dlx3 coded by RNA*iii* show a loss in *dlx4* expression in the visceral arches, indicated by the white arrowhead. **C**, 19% of embryos injected with RNA*iii* lose *dlx4* expression in the otic vesicle (arrowhead). **E**, Over-expression of the Dlx3 truncation encoded by RNA*iii* has no effect on the expression of *dlx4* in the median fin fold (mf) as compared to wild-type embryos, **D**. This is observed in all the over-expression experiments described in this report. All embryos are sagittal views. Symbols as in Fig. 2.2. Scale bar: A, B, C, E, 230 μm; D, 280 μm.

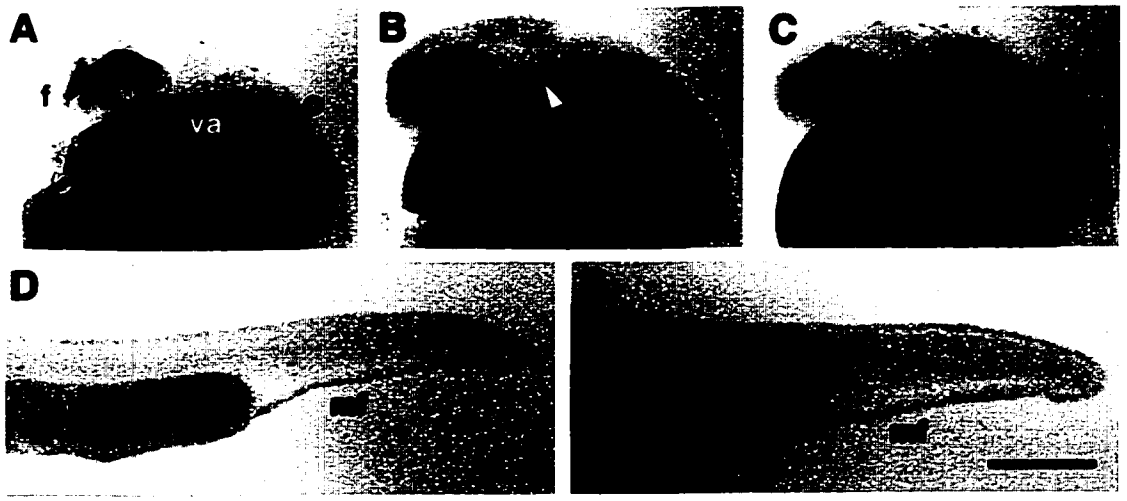

A
f
va
B
C
D

Table 2.1 Summary of perturbations in expression of *dlx4* coincident with overexpression of truncations of Dlx3, as well as a Dlx3-Dlx2 chimera in zebrafish embryos

RNA injected (ng)	Number of embryos examined	*dlx4* expression				Percentage of embryos exhibiting phenotype
		Forebrain	Visceral arches	Ears	Median fin fold	
0		+	+	+	+	
(*i*) 5	50	+	+	+	+	100
(*ii*) 4	27	+	+	+	+	68
		+	+	−	+	15
		+	−	+	+	17
(*iii*)1	105	+	+	+	+	77.5
		+	+	−	+	19
		+	−	+	+	2.5
		−	−	+	+	1
(*iii*) 5	180	+	+	+	+	84
		+	+	−	+	10
		−	−	+	+	3.3
		−	−	−	+	2.7
(*iv*) 4	70	+	+	+	+	63
		+	+	−	+	10
		+	−	−	+	8.5
		−	−	−	+	18.5
(*v*) 4	41	+	+	+	+	75
		+	−	+	+	12.5
		−	−	−	+	12.5

Figure 2.4 Over-expression of a Dlx3-Dlx2 chimera results in perturbations in *dlx4* expression. **A,** Wild-type *dlx4* expression. **B,** Expression of *dlx4* is lost in the visceral arches (arrowhead) of 12.5% of embryos injected with RNA*v*, which codes for the Dlx3-Dlx2 chimeric protein. **C,** In another 12.5% of embryos injected with RNA*v*, *dlx4* expression is lost in the forebrain, visceral arches and otic vesicles (arrowheads). Sagittal views. Symbols as in Fig. 2.2. Scale bar: 100 μm.

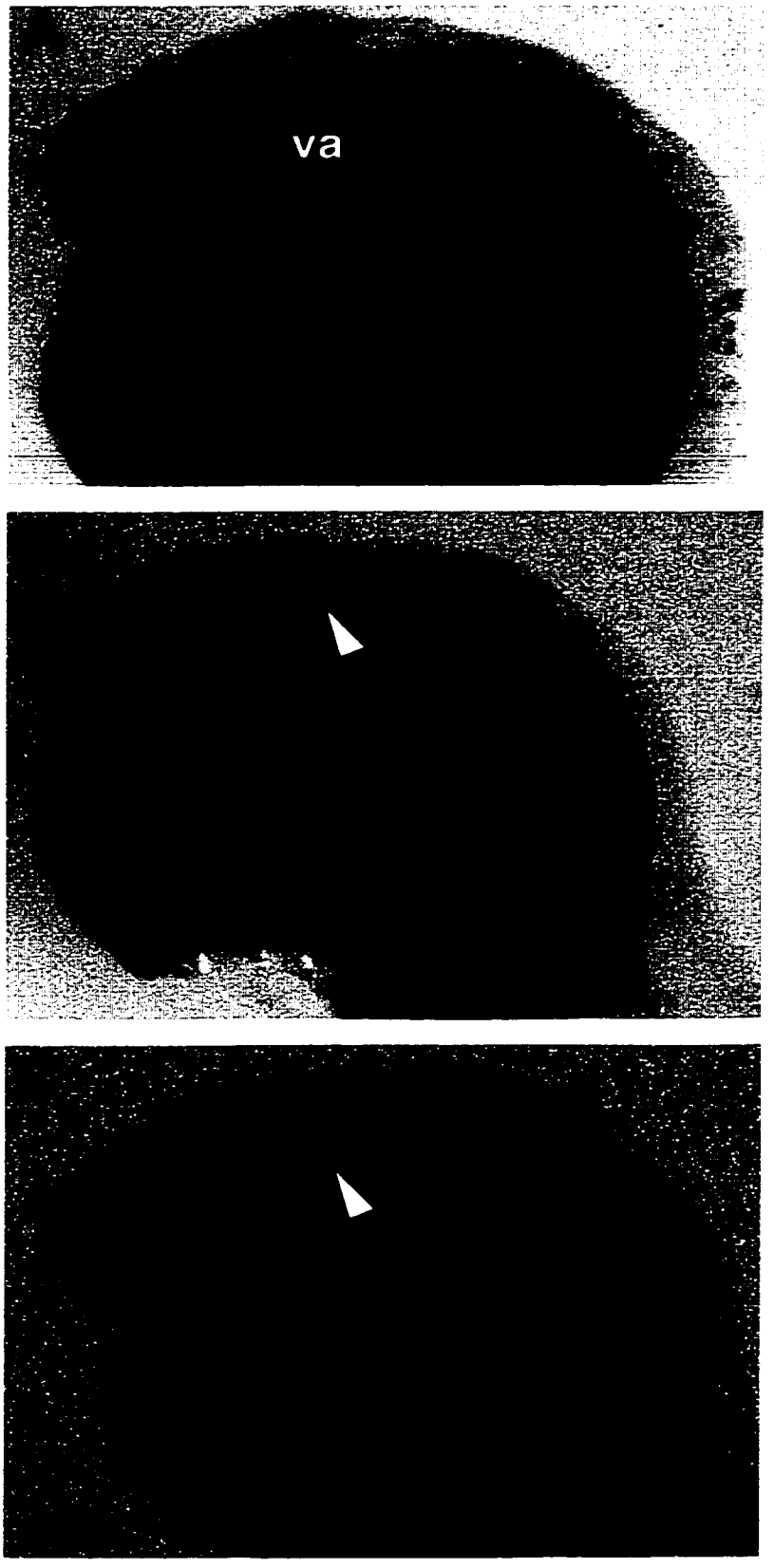
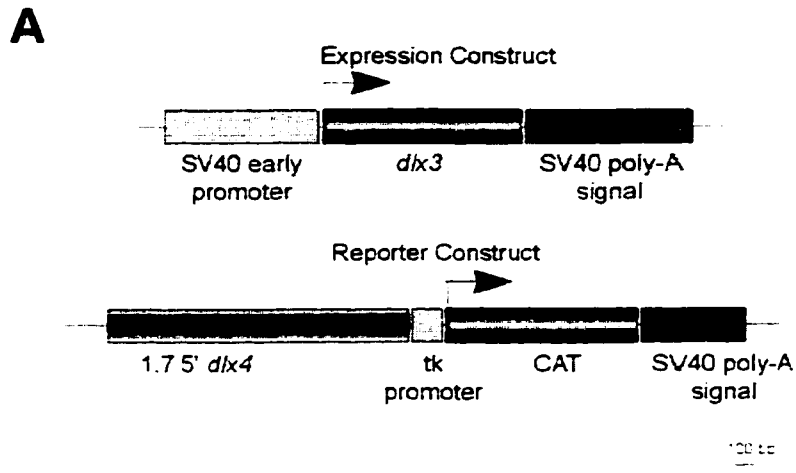

va

Figure 2.5 Dlx3 is able to activate transcription through a 1.7 kb fragment of *dlx4* 5'-flanking sequence. **A,** An expression construct was created by sub-cloning the full length cDNA for *dlx3* downstream of the SV40 early promoter in the pTL2 vector. The reporter consisted of a 1.7 kb DNA fragment located immediately 5' to the *dlx4* gene (1.7 5' *dlx4*), subcloned into the pBLCAT2 vector, upstream of a minimal tk promoter driving expression of the *CAT* gene. The transcriptional start sites of both constructs are indicated by arrows. **B,** Transient co-transfection experiments were performed by calcium-phosphate precipitation in P19 cells to determine the effect of Dlx3 on the 1.7 5' flanking region of *dlx4*. Transcriptional activation through the indicated reporter constructs in the presence or absence of Dlx3. Results represent 3 independent experiments.


A
Expression Construct
SV40 early promoter
dlx3
SV40 poly-A signal
Reporter Construct
1.7 5' dlx4
tk promoter
CAT
SV40 poly-A signal

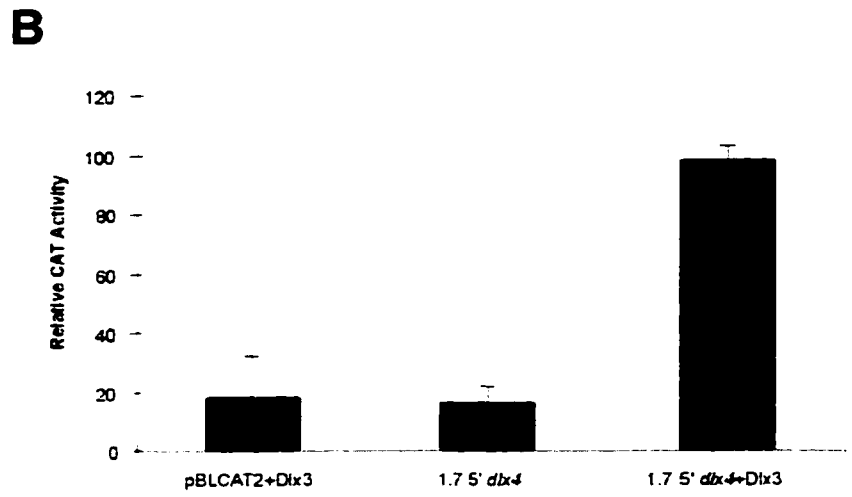
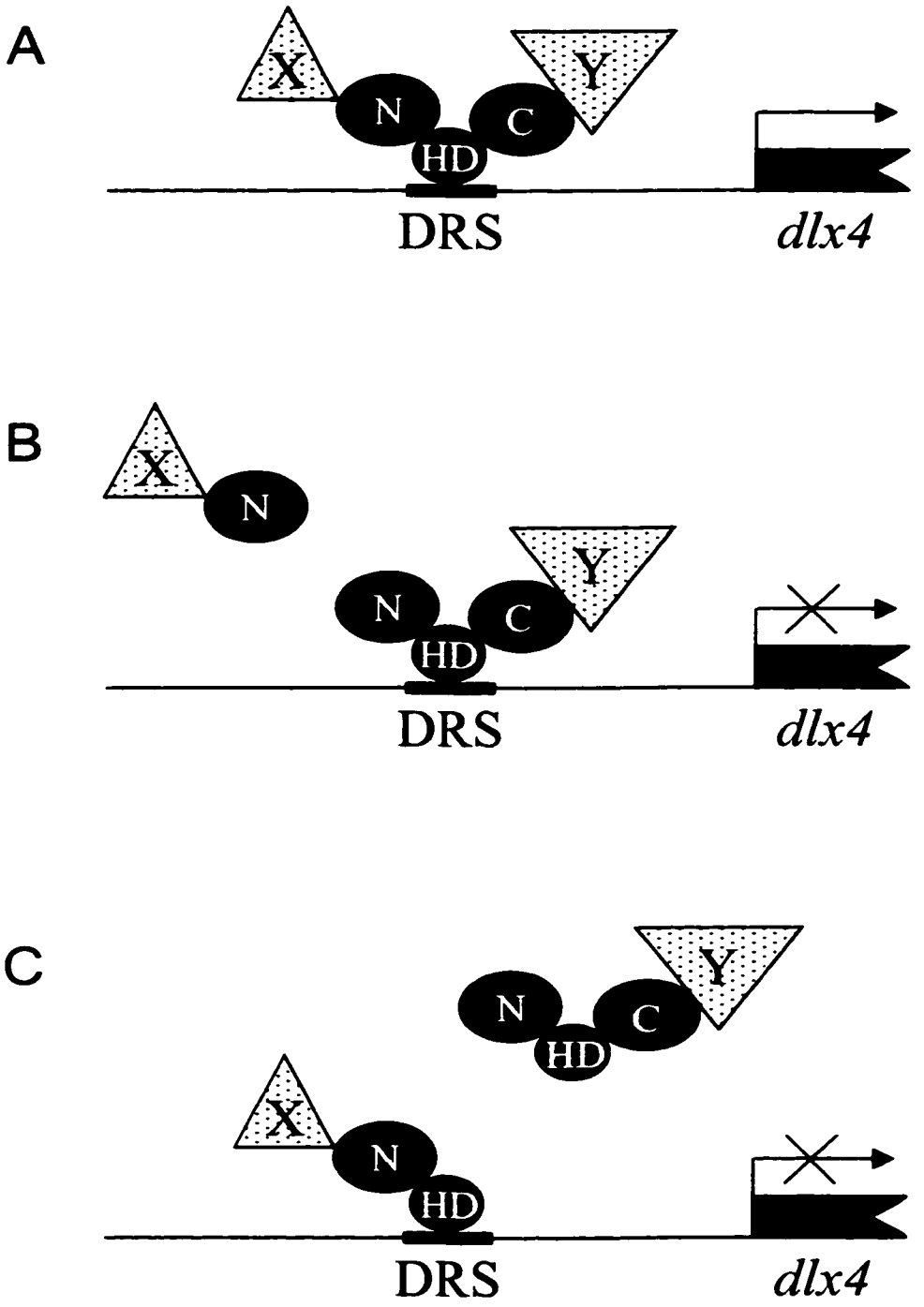

B
Relative CAT Activity
120
100
80
60
40
20
0
pBLCAT2+Dlx3
1.7 5' dlx4
1.7 5' dlx4+Dlx3

Figure 2.6. Proposed model explaining the dominant negative effects of Dlx3 truncations on the activation of *dlx4* mediated by the endogenous Dlx3. **A.** Wild-type transcriptional activation of *dlx4* by Dlx3. The *dlx4* transcription unit is represented by a black box and transcription is indicated by the arrow. The Dlx3 protein is represented as three ovals, each representing one protein domain: N, a putative domain in the N-terminal region; HD, the homeodomain; C, a putative domain in the carboxy-terminal region of the protein. Dlx3 is shown interacting via the homeodomain at its DNA recognition site (DRS) located in the *dlx4* 5'-flanking region. Also shown are two proposed, yet unidentified, co-factors interacting with the Dlx3 protein. The interactions would be required for proper activation of *dlx4* expression. **B.** Squelching of co-factor X by a truncated Dlx3 unable to bind DNA (RNA*iii* or RNA*iv*). This impairs endogenous Dlx3 function at its regulatory sites upstream of *dlx4* by competing for factor X. **C.** A truncated Dlx3 containing the amino terminus and homeodomain (RNA*ii* and, possibly, the Dlx3-Dlx2 chimera) and thus able to bind to DNA, competes for the DRS with the endogenous Dlx3. Unable to interact with factor Y because of the absence of the carboxy-terminal domain, binding of the truncated Dlx3 to the DRS does not result in activation of *dlx4* transcription. An alternative possibility is that interaction with factor Y is necessary for optimal binding of Dlx3 to the DRS. Thus, although containing a homeodomain, the truncated Dlx3 coded by RNA*ii* would act by squelching similarly to what has been proposed for the shorter forms (RNA*iii* or RNA*iv*).


A
X
Y
N
C
HD
DRS
dlx4
B
X
N
Y
N
C
HD
DRS
dlx4
C
Y
N
C
HD
X
N
HD
DRS
dlx4

3. A HIGHLY CONSERVED ENHANCER IN THE *DLX5/DLX6* INTERGENIC REGION IS THE SITE OF CROSS-REGULATORY INTERACTIONS BETWEEN *DLX* GENES IN THE EMBRYONIC FOREBRAIN

T. Zerucha*[1,2], T. Stühmer*[3], B.K. Park[1], Q. Long[1], G. Yu[2], G. Hatch[1], A. Gambarotta[1], J. Schultz[4], J.L.R. Rubenstein[3] and M. Ekker†[1,2,5].

1. Loeb Health Research Institute at the Ottawa Hospital, 725 Parkdale Avenue, Ottawa, On, Canada, K1Y 4E9.
2. Dept. of Cellular and Molecular Medicine, University of Ottawa, Canada
3. Nina Ireland Laboratory of Developmental Neurobiology, Center for Neurobiology and Psychiatry, Department of Psychiatry and Programs in Neuroscience, Developmental Biology and Biomedical Sciences, University of California at San Francisco, 401 Parnassus Avenue, San Francisco, CA 94143-0984
4. Present Address: Tularik Inc., South San Francisco, CA 94080
5. Dept. of Medicine, University of Ottawa, Canada.

* These two authors contributed equally to the work.

† Author for Correspondence. (e-mail: mekker@lri.ca).

Submitted to Development, March, 1999

SUMMARY

Four *Dlx* homeobox genes, *Dlx1*, *Dlx2*, *Dlx5*, and *Dlx6* are expressed in the same primordia of the mouse forebrain with temporally overlapping patterns. The four genes are organized as two tail-to-tail pairs, *Dlx1/Dlx2* and *Dlx5/Dlx6*, a genomic arrangement conserved in distantly related vertebrates like zebrafish. The *Dlx5/Dlx6* intergenic region contains two sequences of a few hundred base pairs, which are remarkably well conserved between mouse and zebrafish. Reporter transgenes containing these two sequences are expressed in the forebrain of transgenic mice and zebrafish with patterns highly similar to endogenous *Dlx5* and *Dlx6* expression. The activity of the transgene is drastically reduced in mouse mutants lacking both *Dlx1* and *Dlx2*, consistent with the decrease in endogenous *Dlx5* and *Dlx6* expression. These results suggest that cross-regulation by Dlx proteins, mediated by the intergenic sequences, is essential for *Dlx5* and *Dlx6* expression in the forebrain. This hypothesis is supported by co-transfection and DNA-protein binding experiments. We propose that the *Dlx* genes are part of a highly conserved developmental pathway that regulates forebrain development.

INTRODUCTION

Genes with regulatory functions are often members of multi-gene families. Their shared structural features allow the prediction that the protein products of these genes will share many biochemical properties. Furthermore, many members of such families are often co-expressed during development. This leads to partial functional redundancy, as evidenced by the weak or lack of phenotype of mutants with targeted mutations affecting single members of gene families. The overlapping expression of multigene family members is likely a direct consequence of a common evolutionary origin and the maintenance of common mechanisms responsible for the control of their expression. However, we propose that when such genes encode for transcriptional regulators, the overlapping expression may also be the result of cross-regulatory interactions between different family members.

The *Dlx* genes of vertebrates constitute a good model to examine such mechanisms. This gene family comprises six members in mammals and at least eight in the zebrafish. Four *Dlx* genes, *Dlx1*, *Dlx2*, *Dlx5*, and *Dlx6* are involved in development of the ventral telencephalon and diencephalon of mammals and the expression patterns of these four genes, although distinct overall, overlap significantly. Mice lacking either *Dlx1* or *Dlx2* function show normal or nearly normal development of the subcortical telencephalon. However, mice lacking both *Dlx1* and *Dlx2* functions show stronger abnormalities in the development of the striatal subventricular zone, in the differentiation of striatal matrix neurons and in the migration of neocortical interneurons from the

subcortical telencephalon (Anderson et al., 1997a,b). Interestingly, expression of *Dlx5* and *Dlx6* is reduced in the subventricular zone, but not in the mantle of the double mutants, suggesting that *Dlx1* and/or *Dlx2* might be required for the maintenance of *Dlx5/Dlx6* expression in subventricular zone cells.

The zebrafish *dlx1*, *dlx2*, *dlx4*, and *dlx6* genes are the orthologs of the mammalian *Dlx1*, *Dlx2*, *Dlx5*, and *Dlx6* genes, respectively. These four zebrafish genes are also expressed in the ventral forebrain with very similar patterns. The similarities between the mouse and zebrafish *Dlx* orthologs also extend to their genomic organization. In both species, the four genes are organized as two pairs of convergently transcribed genes, the *Dlx1/Dlx2* pair and the *Dlx5/Dlx6* pair (*dlx4/dlx6* in zebrafish). The relatively short distances (2.5 to 10 kb) that separate the two genes in such pairs and the similarities in the expression patterns of the two genes that constitute a pair (Ellies et al., 1997a,b) suggest the presence, in the intergenic region, of shared *cis*-acting regulatory elements.

In the present study, we have examined the molecular basis for the overlapping expression of *Dlx* genes in the ventral forebrain of vertebrates. We have identified highly conserved sequences in the intergenic region between *Dlx5/Dlx6 (dlx4/dlx6)*. These sequences extend over a few hundred base pairs and are the potential site of action of a vast number of regulatory factors. We present evidence that the Dlx proteins themselves constitute some of these factors suggesting that cross-regulatory mechanisms between *Dlx* genes is an important aspect of their regulation in the forebrain.

MATERIALS AND METHODS

Identification of conserved sequences in the zebrafish *dlx4/dlx6* and mouse *Dlx5/Dlx6* intergenic regions

Restriction fragments of a genomic clone containing the zebrafish *dlx4/dlx6* locus (Ellies et al., 1997b) were radiolabeled and hybridized to a Southern blot of various restriction digests of a mouse genomic clone containing the orthologous *Dlx5/Dlx6* locus (Liu et al., 1997). Of the zebrafish restriction fragments from the *dlx4/dlx6* locus, only a 1.4 kb *Xho*I-*Eco*RI fragment from the intergenic region hybridized to the mouse genomic fragments (Fig. 3.1). This zebrafish fragment and the hybridizing mouse fragments were sequenced using the dideoxy procedure. A search of the GenBank database with the zebrafish 1.4 kb *Xho*I-*Eco*R1 fragment enabled us to identify a human BAC clone containing the *DLX5/DLX6* locus (sequence accession number AC004774).

Nucleotide sequence comparisons were done using the GCG software package and the CLUSTAL W version 1.7 (Thompson, 1984) multiple sequence alignment program.

Production of transgenic animals

DNA fragments from either the zebrafish *dlx4/dlx6* locus or from the mouse *Dlx5/Dlx6* locus were subcloned into the p1229 or p1230 vectors (Yee and Rigby, 1993). For the production of transgenic mice, the transgene was excised from the plasmid construct and injected at a concentration of 5 ng/μl in eggs from FVB/n crosses using standard procedures (Hogan et al., 1986). Transgenes were analyzed in either founder embryos or from established transgenic lines. Presence of the transgene was determined by PCR on DNA prepared from extra-embryonic tissues with the following oligonucleotide primers 5'-AGGGCAGAGCCATCTATTGC-3' and 5'-CGCTCATCCGCCACATATCC-3' derived, respectively, from the β-*globin* promoter and *lacZ* sequences of the p1229/p1230 vectors. Amplification of a fetal hemoglobin gene sequence was used as a positive control (primers are x1: 5'-GATCAT GACCGCCGTAGG-3' and x2: 5'-CATGAACTTGTCCCAGGCTT-3'.

For the production of transgenic zebrafish, a 1.4 kb *Xho*I-*Eco*RI fragment of the zebrafish *dlx4/dlx6* intergenic region was inserted upstream of the β-*globin* promoter fragment taken from the p1230 vector and of the coding sequence of a variant of the green fluorescent protein, GM2, that emits approximately 30-fold higher fluorescence than does the wild-type green fluorescent protein (GFP), under standard FITC conditions (Cormack et al., 1996). Linearized plasmid DNA was injected into single-cell wild-type zebrafish embryos which were examined for GFP expression at various time points thereafter as previously described (Long et al., 1997).

Morphological analysis of transgenic animals

Founder transgenic embryos or embryos from the cross of a transgenic male with normal *FVB* or *CD1* females were harvested at various embryonic stages. Transgene expression was also analyzed in newborn pups, young mice and adults from established lines. Embryos were fixed in 1% formaldehyde; 0.2% glutaraldehyde; 0.02% NP-40 in phosphate-buffered saline (PBS) for 30 min at 4° C, washed in PBS for 20 min at room temperature and stained for β-galactosidase activity overnight at 28° C in a solution of 1 mg/ml X-gal, 5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6$, 2 mM $MgCl_2$, 0.02% NP-40 in PBS.

Breeding with mouse null mutants.

Mice heterozygous for the *zfdlx4/6lacZ*-transgene were mated to partners heterozygous for a deletion of the *Dlx1* and *Dlx2* genes (strain C57 Bl/6; described in (Qiu et al., 1997). Offspring was genotyped by PCR, with the primers described above to detect the transgene, and with primers corresponding to the neomycin resistance gene that marks the *Dlx1/2* deletion. Animals that were heterozygous for both modifications (genotype *Dlx1/2+/-*; *zfdlx4/6lacZ+/-*) were mated to generate mice that were homozygous for the deletion of the *Dlx1/2* locus (*Dlx1/2-/-*; *zfdlx4/6lacZ*). Mutant embryos were identified by either diagnosing a cleft palate (E15 and later) or by the absence of a PCR product, with primers that recognize the deleted portion of *Dlx1*.

Constructs for transient co-transfection experiments

An effector plasmid that expresses the zebrafish *dlx2* gene under control of the SV40 early promoter was constructed by inserting an 845 bp PCR-amplified *Eco*RI fragment of the zebrafish *dlx2* cDNA (Akimenko et al., 1994) encompassing the full coding sequence into the *Eco*RI site of the pTL2 expression vector (M. Petkovich, unpublished). Reporter plasmids were constructed by inserting fragments of the zebrafish *dlx4/dlx6* intergenic region into the pBLCAT2 vector (Luckow and Schütz, 1987) which contains the thymidine kinase (tk) minimal promoter driving expression of the chloramphenicol acetyltransferase (CAT) gene. The 1.4 kb *Xho*I/*Eco*RI fragment from the zebrafish *dlx4/dlx6* intergenic region (Id4/d6 Fig. 3.1) was subcloned into pBLCAT2 directly upstream of the *tk* promoter. The Id4/d6i and Id4/d6ii fragments and deletions of Id4/d6i were prepared by PCR from a pBluescript clone containing the 1.4 kb *Xho*I/*Eco*RI Id4/d6 fragment. The Id4/d6ii fragment was inserted in pBLCAT2 directly upstream of the *tk* promoter and the Id4/d6i fragment was inserted immediately downstream of the CAT gene (i.e. 4.5 kb upstream of the *tk* promoter in the circular plasmid).

The following oligonucleotides:

1060: 5'-GCTCTAGAATTAGTTTAACGTCGAA-3';

473: 5'-GGGGTACCGCTGGGGCATCCACGAT-3';

187: 5'-GGGGTACCATTCTCATAAATGCAG-3';

204: 5'-GGGGTACCTGCATTTATGAGAATG-3';

305: 5'-GGGGTACCATCTTTATTTGGATT-3';

316: 5'-GGGGTACCAAATAAAGATGCCTTT-3' were used to prepare deletion fragments from Id4/d6i using PCR. The numeric name of the oligonucleotide refers to the position, in the conserved intergenic sequence (Fig. 3.2A), that borders the amplified fragment. Restriction sites (*Xba*I or *Kpn*I) were introduced at the 5'-end of each oligonucleotide. PCR products: full length Id4/d6i (positions 1-473 in sequence shown in Fig. 3.2A), Id4/d6i 1-204, positions 1-204 in the same sequence; Id4/d6i 187/316, and Id4/d6i 305/473, were PCR amplified and subcloned into pBLCAT2.

Transient co-transfection experiments

Transient co-transfection experiments were performed in the P19 murine embryonic carcinoma (EC) cell line essentially as described previously (Zerucha et al., 1997). Cells were seeded 24 h prior to transfection at a density of 10^7 cells per 100 mm dish. Transfections were carried out by the calcium phosphate precipitation procedure (Sambrook et al., 1989). A total of 10 μg of DNA per dish was used in each transfection. This included: 2 μg pRSV-βgal as an internal control for transfection efficiency, 2 μg of reporter plasmid, 2 μg of effector plasmid, and sheared calf thymus DNA (Boehringer Mannheim) to the total of 10 μg. Precipitates were left on the cells for 16 h and the cells harvested 64 h post-transfection. Cells were collected in PBS, pelleted by centrifugation and resuspended in freeze/thaw buffer (250 mM Tris-HCl pH 8, 10 mM DTT, 15% glycerol). Cell extracts were prepared by repeated cycles of freezing/thawing. β-Galactosidase activity was assayed as described by Sambrook et al. (Sambrook et al., 1989). CAT activity was determined by thin-layer chromatography and measured as

percent conversion of mono- and diacetylated chloramphenicol relative to unmodified plus acetylated chloramphenicol using the BIO-RAD GS-525 Molecular Imager System. CAT activity was standardized to β-galactosidase levels to compensate for variations in transfection efficiency. Experiments were performed in duplicate and repeated a minimum of three times. Error bars in the figures represent standard errors on the means of all replications.

Stable transfectant cell line

A PCR-amplified *Eco*RI fragment encompassing the full length (845 bp) coding region of zebrafish *dlx2* cDNA (described above) was subcloned into the pTL-MTG vector (Prefontaine et al., 1998) downstream of and in frame with 6 repeats of a c-*myc* sequence which encodes a polypeptide consisting of an epitope recognized by the 9E-10 monoclonal antibody (*myc*-tag; Santa Cruz Biotech). Expression of this fusion protein is under control of the SV40 early promoter. This construct (pTL-MTG-Dlx2) was co-transfected together with pCMVneo into SF7 *SCID* fibroblastic cells as described above using the calcium phosphate procedure with the following modifications: 8 μg pTL-MTG-Dlx2 and 2 μg pCMVneo made up the total DNA transfected per 100 mm dish; 40 h post-transfection 600 μg/ml G418 was added to the cells. Cells were maintained in this concentration of G418 until the formation of discernable colonies. Individual colonies of cells were isolated and grown separately. Each clone was screened by PCR for the presence of a zebrafish *dlx2* sequence. MTG-Dlx2 protein was prepared from nuclear extracts of the stable transfectant cell line SF7-MTG-Dlx2 essentially as described by

Andrews and Faller (Andrews and Faller, 1991). In brief, cells from each confluent 100 mm dish were harvested and resuspended in 1.5 ml PBS on ice. The cell suspension was pelleted and resuspended in 400 µl of a cold solution of 10 mM HEPES-KOH pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 0.2 mM PMSF, incubated on ice 10 min, then vortexted 10 sec. Insoluble nuclei were pelleted and the supernatant, containing cytoplasmic contents and outer membrane, was discarded. Nuclei were resuspended in 20-100 µl of a cold solution of 20 mM HEPES-KOH pH 7.9, 1.5 mM $MgCl_2$, 420 mM NaCl, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSF, 25% glycerol, and incubated on ice 20 min. Cellular debris was removed by centrifugation. Protein concentration of the supernatant was determined by the BIO-RAD protein assay, and single use aliquots stored at -80° C.

The presence of MTG-Dlx2 was determined by immunoblotting with the 9E-10 monoclonal antibody. Four individual clones were positive for MTG-Dlx2 after both rounds of screening and one was chosen for subsequent experiments. It was thereafter maintained in 400 µg/ml G418.

Electrophoretic mobility shift assays

A 130 bp DNA fragment corresponding to positions 187-316 of Id4/d6i (Fig. 3.2A) was amplified by PCR using oligonucleotides 187 and 316. The PCR product was inserted into the pCRII vector (Invitrogen). The fragment was excised with *Eco*RI prior to filling of the 5' overhangs with the large fragment of DNA polymerase I (Klenow) in

presence of radiolabelled [α^{32}P]dATP. Binding reactions were carried out in a total volume of 20 µl in 12 mM HEPES-KOH (pH 7.9), 1 mM EDTA, 0.4 mM $MgCl_2$, 100 mM NaCl, 0.6 mM DTT, 0.6 mM PMSF, 13% glycerol. Nuclear extract (10 µg) from the stable SF7-MTG-Dlx2 cell line or from control SF7 cells was pre-incubated with the 9E-10 anti-myc monoclonal antibody or an equivalent volume of water at room temperature for 30 min. Following pre-incubation, 1 µg BSA, 1 µg sheared calf-thymus (Boehringer Mannheim) and 15,000 cpm of radiolabelled probe were added and incubated at room temperature for 20 min. Protein-DNA complexes were resolved on a 4% polyacrylamide (29:1 acrylamide:bis-acrylamide) gel run in 1x TBE.

RESULTS

Identification of highly conserved elements in the zebrafish *dlx4/dlx6* and mouse *Dlx5/Dlx6* intergenic regions.

We identified two highly conserved sequences in the region between the zebrafish *dlx4* and *dlx6* genes and their mouse orthologs *Dlx5/Dlx6*. A 1.4 kb *Xho*I/*Eco*RI fragment from the zebrafish *dlx4/dlx6* intergenic region (Fig. 3.1) was found to hybridize to a pair of restriction fragments in the mouse *Dlx5/Dlx6* intergenic region. Nucleotide sequence analysis revealed two sequences, of about 400 bp and 300 bp, respectively, that are highly similar between the two species (Fig. 3.2 A, B). The orientation and the relative position of the two sequences relative to *dlx4* (*Dlx5*) and *dlx6* (*Dlx6*) are identical (Fig. 3.1). The 400 bp sequence, named Id4/d6i is closer to the *dlx6* gene than is the 300

bp sequence, hereafter named Id4/d6ii. The orthologous mammalian elements are hereafter called ID5/D6i and ID5/D6ii, respectively. We have identified highly similar sequences at the human *DLX5/DLX6* locus by searching the GenBank database (Fig. 3.2A, B).

Nucleotide sequence comparisons indicate the human and mouse ID5/D6i elements to be identical except for a 3-bp insertion in the human sequence (Fig. 3.2A). The zebrafish Id4/d6i sequence is 83% identical to its mammalian counterparts over 384 bp (Fig. 3.2A), including a central 131 bp with 94% sequence identity. The human and mouse ID5/D6ii sequences are 98% identical and zebrafish Id4/d6ii shares about 85% identity over 275 bp with its mammalian counterparts (Fig. 3.2B). It lacks a stretch of adenines, about 20 nucleotides in length, that is found in the two mammalian sequences.

A zebrafish *dlx4/dlx6* intergenic fragment targets reporter gene expression to the forebrain and olfactory placodes in transgenic mice

A construct containing the *lacZ* reporter gene under the control of a β-globin minimal promoter and the entire zebrafish *dlx4/dlx6* intergenic region (plus a short segment of the transcription unit of *dlx4* and a few base pairs of the transcription unit of *dlx6*; *zfdlx4/6lacZ* transgene; Fig. 3.1) was injected into fertilized mouse eggs to produce transgenic animals. We obtained seven lines of transgenic mice. All seven lines showed *lacZ* expression, beginning around E10, in two groups of forebrain cells, one in the basal telencephalon and one in the ventral thalamus/hypothalamus (indicated as I and II in Fig.

3.3A). Examination of whole-mount embryos stained for β-galactosidase activity indicated that the patterns of reporter transgene expression are strikingly similar to patterns of endogenous mouse *Dlx* expression in the forebrain (Shimamura et al., 1997). Mouse embryos express *Dlx5* and *Dlx6* in two separate domains within the forebrain. Domain I is a longitudinal alar plate stripe that begins at the zona limitans intrathalamica and extends rostrally through the ventral thalamus (VT) and several hypothalamic areas (Hy) to the rostral midline. Domain II is a longitudinal region in the basal telencephalon that extends rostrally from part of the caudal ganglionic eminence (CGE, amygdala primordium), through the lateral and medial ganglionic eminences (LGE and MGE), and into the septal and preoptic (POA) primordia (Bulfone et al., 1993a,b; Puelles et al., manuscript in preparation).

Activity of the β-galactosidase reporter transgene persisted in the forebrain of transgenic mice until adulthood (Fig. 3.3B, C and data not shown). In addition to the developing forebrain, the reporter transgene under the control of zebrafish sequences was expressed in the olfactory placodes in 5/7 mouse lines (Fig. 1; Fig. 3.3A, B). There were very few additional sites of transgene expression: one line had a few labeled eye cells; one line had expression in the developing shoulder area and one showed expression in the trunk somites (data not shown). Additional sites of endogenous mouse *Dlx5* and *Dlx6* expression were negative, including the branchial arches, the otic vesicle and the limb apical ectodermal ridge.

To assess the degree to which expression of the *zfdlx4/6lacZ* transgene matches endogenous *Dlx5* and *Dlx6* expression in the forebrain, we have compared their expression patterns, using X-gal staining and radioactive *in situ* hybridization, respectively, on transverse brain sections. We analyzed sequential sections from E10.5, E12.5, E14.5, E17.5, and P0 mice.

Dlx5 and *6* are expressed in Domains I and II in slightly different, but overlapping patterns: *Dlx5* is expressed strongest in the subventricular zone (SVZ) whereas *Dlx6* is expressed strongest in the mantle zone (MZ) (Fig. 3.3E,F). Neither gene is expressed appreciably in the ventricular zone (VZ) (Liu et al., 1997).

Remarkably, the *zfdlx4/6lacZ* transgene is expressed in a pattern that is extremely similar to the mouse *Dlx5* and *6* genes. β-Galactosidase activity and *Dlx5* transcripts can first be detected in the forebrain at ~E10.5, as a thin layer of cells overlying the ventricular zone in the basal telencephalon. On E10.5, E12.5 and E14.5, zebrafish *dlx4/6*-enhancer driven β-galactosidase expression in the mouse telencephalon matches mouse *Dlx5* expression (Fig. 3.3D, E, G, H and data not shown). In Domain I, the *zfdlx4/6lacZ* transgene has a pattern that appears to be intermediate between *Dlx5* and *Dlx6*; note that it lacks the strong SVZ expression of *Dlx5* (Fig. 3.3G, H).

The zebrafish *dlx4/dlx6* intergenic enhancer is active in the forebrain of zebrafish embryos.

To determine if the intergenic sequences that target reporter gene expression to the forebrain of transgenic mice can reproduce *dlx* expression in zebrafish embryos, we microinjected, into one cell stage embryos, a construct that contained the 1.4 kb zebrafish *Eco*RI-*Xho*I intergenic fragment from *dlx4/dlx6* (Fig. 3.1), the same β-globin minimal promoter fragment as for the transgenic mouse experiments, and the gene coding for the green fluorescent protein (GFP) as a reporter. Primary transgenic zebrafish embryos carrying this construct expressed GFP specifically in forebrain cells forming two domains (Fig. 3.3J) with patterns strikingly similar to endogenous *dlx4/dlx6* expression (Fig. 3.3K, L). Of the 750 embryos that survived microinjection until the second day of embryonic development, 4 had very high levels of GFP expression in the forebrain, 25 had 5-10 GFP-positive forebrain cells, 30 had 1 or 2 positive forebrain cells. Fifteen embryos showed one or a few GFP-positive cells in ectopic locations. The onset of GFP expression in the forebrain was around 17-19 hours after fertilization (h.p.f.), shortly after the onset of *dlx4* expression as detected by *in situ* hybridization. GFP expression persisted in the forebrain until at least 36 h.p.f.

The two domains of *dlx* expression in the forebrain of zebrafish embryos (Akimenko et al., 1994) are reminiscent of the two domains observed in the mouse embryonic forebrain. To compare expression patterns of *dlx* genes in the zebrafish forebrain, we made sections of 48 h-old embryos hybridized with *dlx* probes.

Interestingly, the patterns of expression of *dlx1* and *dlx2* in both the telencephalon and the diencephalon indicate that the two genes are expressed in more immature cells, as reflected by their position closer to the ventricular walls than the cells that express *dlx4* and *dlx6* (Fig. 3.4). A similar observation had been made previously for the mouse orthologs of these four genes (Liu et al., 1997).

Most of the forebrain activity of the *dlx4/dlx6* intergenic enhancer is located in Id4/d6i.

The two conserved sequences located in the zebrafish *dlx4/dlx6* intergenic region (Id4/d6i and Id4/d6ii) were inserted separately into reporter constructs and used to produce transgenic mouse embryos. At E11, forebrain *lacZ* expression targeted by the Id4/d6i enhancer construct was indistinguishable from the full-length Id4/d6 enhancer (compare Fig. 3.5A with Fig. 3.3B) although none of the embryos showed expression in the olfactory epithelium (n = 12; Fig. 3.1).

We also generated transgenic mice with a reporter construct that contained the second conserved sequence from the zebrafish *dlx4/dlx6* intergenic region, Id4/d6ii. Two lines of transgenic mice and ten primary transgenic mouse embryos were produced. Embryos from one transgenic line showed *lacZ* expression in the olfactory epithelium (data not shown) resembling that obtained with the full-length Id4/d6. This was also observed in one primary transgenic embryo, which, in addition, had expression in the apical ectodermal ridge of the limb buds, where *Dlx* genes are expressed (Bulfone et al.,

1993; Dollé et al., 1992). Finally, one primary Id4/d6ii transgenic embryo showed correct *lacZ* expression in the forebrain (data not shown). Thus, Id4/d6ii was much less efficient at targeting *lacZ* to the forebrain (0/2 lines; 1/10 primary transgenic embryos) compared to full length Id4/d6 (7/7 lines) or to Id4/d6i (12/12 primary transgenic embryos; Fig. 3.1).

We next tested whether the orthologous mouse ID5/D6i could regulate correct *Dlx* expression in the forebrain. Four stable lines and 4 primary transgenic embryos were produced with the mouse ID5/D6i construct and nearly all showed forebrain expression (Fig. 3.5B); one transgenic line did not express *LacZ* anywhere, possibly due to an integration effect. Reversing the orientation of mouse ID5/D6i had no effect on its expression (9/9 primary transgenic embryos; Fig. 3.1 and data not shown). Thus, both the orthologous mouse ID5/D6i and the zebrafish Id4/d6i fragments are enhancers that efficiently replicate the correct pattern of *Dlx* expression in the forebrain.

Unlike the zebrafish Id4/d6i enhancer, mouse ID5/D6i in either orientation frequently reproduced correct *Dlx* expression in the branchial arches (2/4 stable lines and 7/13 primary embryos (sum of both orientations, Fig. 3.1, 3.5B and data not shown), olfactory placode (1 line and 1 primary embryo) and apical ectodermal ridge (1 line; Fig. 3.1). No expression in the otic vesicle was observed in any embryos but some expression was detected in the middle ear that is consistent with the branchial arch expression.

To begin to identify the essential sequences within these enhancers, we used a deletion fragment of the zebrafish Id4/d6i enhancer, corresponding to positions 187-316 (Fig. 3.1 top, 3.2A), and examined its activity in transgenic mouse embryos at E11. Of five primary transgenic embryos, all appeared to have correct expression in Domain II in the forebrain (Fig. 3.5C). However, β-galactosidase expression in Domain I (ventral thalamus and hypothalamus) was occasionally weaker or not detectable (data not shown). A similar construct also targeted expression of GFP principally to the forebrain of zebrafish embryos (data not shown).

Activity of the zebrafish intergenic enhancer is reduced in mice lacking *Dlx1* and *Dlx2*

Mutant mice that lack both *Dlx1* and *Dlx2* function have a time-dependent block in basal telencephalon differentiation (Anderson et al., 1997a,b). While early neurogenesis appears to be normal, later neurogenesis is not. This phenotype seems to be due to a defect in the production and/or function of the subventricular zone. Accordingly, in *Dlx1/2* mutants *Dlx5* and *Dlx6* expression is not detectable in the subventricular zone of the LGE and MGE, but is maintained in early born mantle cells at E12.5 (Anderson et al., 1997a,b). As described above, the zebrafish and mouse intergenic enhancers are highly active in the SVZ of the basal telencephalon. Therefore, it is possible that the Dlx1 or Dlx2 proteins might be, at least in part, responsible for the activity of this enhancer. To test this hypothesis, we bred mice containing the zebrafish *dlx4/dlx6* full intergenic reporter construct with mice heterozygous for a mutation that inactivates both

Dlx1 and *Dlx2*. We then inbred mice that are heterozygous for both the mutation and the transgene to generate *Dlx1/2-/-* homozygotes that also had at least one *zfId4/d6lacZ* allele.

In embryos that are homozygous for the *Dlx1/Dlx2* mutation, β-galactosidase activity is strikingly reduced in the subventricular zone of the developing striatum (Fig. 3.6A-D). These results parallel the changes in endogenous *Dlx5* and *Dlx6* expression in the *Dlx1/2*$^{-/-}$ mutant mice (Fig. 3.6E-L). Based on these results, we propose that *Dlx1* and/or *Dlx2* function is required, directly or indirectly, to regulate *Dlx5* and *Dlx6* expression via their intergenic enhancer.

To determine whether the loss of *Dlx5*, *Dlx6* and *zfDlx4/6lacZ* expression in the SVZ of the *Dlx1/2* mutants is due to a loss of those cells or to a change in gene regulation in SVZ cells, we studied the expression of *Dlx1* and *Dlx2* mRNAs. In the *Dlx1/2* mutants, the 5' end of these genes was not deleted (Qiu et al., 1997). Thus, if the *Dlx1/2* *cis*-acting regulatory sequences responsible for *Dlx1/2* expression are intact, and the truncated *Dlx1* and/or *Dlx2* transcripts are stable, we should be able to detect the cells that normally express *Dlx1/2* in these mutants. In fact, *in situ* hybridization demonstrates that both truncated genes are still expressed in the proliferative zones of the *Dlx1/2* mutants (Fig. 3.6M-P and data not shown). This strongly supports the model that cells expressing *Dlx1/2* are maintained in the mutants, and that there is molecular dysregulation within these cells leading to the loss of *Dlx5*, *Dlx6* and *zfdlx4/6lacZ* expression.

Dlx proteins can up-regulate transcription from conserved intergenic sequences.

Dlx proteins bind DNA and can regulate transcription (Liu et al., 1997; Zhang et al., 1997). Because analysis of *Dlx1/2*$^{-/-}$ mutant mice suggest that Dlx1 and/or Dlx2 function is necessary for proper expression of *Dlx5/Dlx6*, one possibility is that this interaction is directly mediated by transcriptional activation of the intergenic enhancer(s) by Dlx1 or Dlx2. To test this model, we performed transient co-transfection assays in cultured cells. Reporter plasmids were constructed to contain either the zebrafish 1.4 kb Id4/d6 fragment, which contains both Id4/d6i and Id4/d6ii, or to contain only one of these elements. Effector plasmids were constructed to express full length zebrafish Dlx1, Dlx2, Dlx3, Dlx4, or Dlx6 proteins under the control of the SV40 early promoter, or full-length mouse Dlx1, Dlx2, or Dlx5 proteins under the control of the CMV promoter.

Co-transfection, into mouse P19 murine embryonic-carcinoma cells, of a construct expressing the zebrafish Dlx2 protein resulted in a 20-fold increase in the activity of the CAT reporter gene placed under the control of the 1.4 kb Id4/d6 intergenic fragment (Fig. 3.7A). All of the zebrafish *Dlx* expression vectors were able to activate expression of the reporter construct to a similar extent (data not shown). The mouse Dlx1, Dlx2, and Dlx5 expression vectors were also able to activate transcription of the same reporter construct in a neuroepithelial cell line (MNS-71; Yu et al., 1999). This indicates that Dlx proteins from either zebrafish or mouse are able to recognize similarly the zebrafish intergenic enhancer sequences in at least two different cell types. However, it does not seem that this effect can be obtained in every cell type because a stimulation

of reporter gene activity by the zebrafish or by the mouse Dlx proteins could not be seen in NIH 3T3 cells (data not shown). Not all homeodomain proteins could activate transcription from the 1.4 kb Id4/d6 intergenic fragment; the products of *sine oculis*-related genes (*six* genes) were unable to activate transcription from this sequence (data not shown).

To determine if both Id4/d6i and Id4/d6ii contain targets for Dlx proteins, the same *Dlx* expression constructs were co-transfected with reporter constructs containing either Id4/d6i or Id4/d6ii. All Dlx proteins examined activated expression through the Id4/d6i element (Fig. 3.7A and data not shown) but none activated transcription from a reporter containing Id4/d6ii (Fig. 3.7A and data not shown), except perhaps for a weak (less than two-fold activation) by the mouse Dlx5 protein. Furthermore the degree of activation produced by Dlx proteins on Id4/d6i reporter constructs were comparable to those obtained with the 1.4 kb Id4/d6 fragment.

In an attempt to narrow down the region of Id4/d6i required for activation by Dlx proteins, a series of deletion fragments of Id4/d6i were prepared and subcloned into the reporter plasmid. The orientation of each of the deletion fragments was maintained relative to the orientation of the full-length Id4/d6i. In transient co-transfection experiments, Dlx2 activated transcription of constructs containing the 187-316 deletion fragment, to an extent similar to that observed with the full-length Id4/d6i (Fig. 3.7A). However, Dlx2 did not activate transcription of constructs containing either the 1-204 or 305-450 fragments (Fig. 3.7A).

To determine if the Dlx proteins are able to directly interact with the Id4/d6i element, electrophoretic mobility shift assays (EMSA) were performed. Nuclear extracts from SF7 *SCID* fibroblasts expressing a fusion of a *c-myc* fragment with full-length zebrafish Dlx2 (MTG-Dlx2) produced a lower mobility complex with the 187-316 deletion fragment of Id4/d6i (Fig. 3.7B). Migration of the lower mobility complex obtained with MTG-Dlx2 was further retarded in the presence of an anti-c-myc antibody, indicating the lower mobility complex contains MTG-Dlx2. No complexes of lower mobility were obtained with a control SF7 nuclear extract (Fig. 3.7B). Furthermore, neither of the other two deletion fragments of Id4/d6i produced a complex of lower mobility in the presence of MTG-Dlx2-containing SF7 extract (Fig. 3.7B), a result consistent with the absence of activation, by Dlx2, of reporter constructs containing these intergenic fragments in co-transfection experiments.

DISCUSSION

One intergenic enhancer is sufficient to recapitulate forebrain *Dlx* expression

A zebrafish sequence from the intergenic region between the *dlx4* and *dlx6* genes is sufficient, once combined to a minimal promoter, to direct expression in those cells of the telencephalon and diencephalon, of either zebrafish or mice, that normally express *Dlx* genes (Fig. 3.3). This strongly suggests that the regulatory mechanisms controlling *Dlx* expression in the forebrain have been conserved during vertebrate evolution and it lends support to the idea that *Dlx* function during forebrain development has also been

conserved. Additional evidence for conserved function of *Dlx* genes in forebrain development comes from the differential expression of *Dlx* genes in the telencephalon and diencephalon where more immature cells express *Dlx1* and *Dlx2* compared to *Dlx5* and *Dlx6*, as seen both for the mouse genes (Liu et al., 1997) and for their zebrafish orthologs (Fig. 3.4 and Akimenko et al., 1994).

Functional conservation of enhancer sequences between mammals and teleost fish has been previously observed for the *Otx2* (Kimura et al., 1997), *hoxb1* (Marshall et al., 1994), and the *Hoxd-11* (Beckers et al., 1996; Gérard et al., 1997), although in the latter case, temporal, spatial and mechanistic differences could be observed between the fish enhancer and its mammalian counterpart.

Because the two *Dlx* genes that constitute a linked pair on a chromosome are organized in an inverted and convergent manner, the intergenic enhancer we have identified is located downstream of the transcription units of both genes. Consistent to known properties of enhancer sequences, the full Id4/d6 and Id4/d6i sequences are able to function when placed upstream of a minimal promoter in a linear construct, and function similarly regardless of their orientation relative to the promoter.

Comparisons of the enhancer activities of Id4/d6i (ID5/D6i) and Id4/d6ii suggest the former plays a more important role in forebrain expression. Id4/d6ii may still be necessary for optimal *Dlx* expression in the ventral forebrain but this enhancer may

require, in order to function efficiently, the presence of other regulatory sites, either from Id4/d6i or from the promoters of one or both flanking genes.

Detailed analysis of reporter transgene activity in the mouse forebrain suggests that both the full dlx4/dlx6 intergenic construct and the Id4/d6i (ID5/D6i) sequences reproduce the endogenous *Dlx5* expression pattern more faithfully than the *Dlx6* expression pattern. This was observed principally by comparison of β-galactosidase expression and endogenous transcript levels in the LGE and MGE of the telencephalon (Fig. 3.3). In other areas, such as the ventral thalamus and hypothalamus, the pattern of β-galactosidase activity was intermediate between those for *Dlx5* and *Dlx6*. Identical results were obtained with constructs from either zebrafish or mouse origin. Therefore, the observed differences cannot be attributed solely to the inability of a zebrafish enhancer to precisely recapitulate *Dlx* expression in a mouse embryo. One possible explanation for this result is that sequences, necessary for maximal expression in cells of the mantle, are absent from our constructs, and, therefore, located outside the intergenic region. It is also possible that, although the intergenic enhancer is sufficient to direct expression to the ventral forebrain, its activity is modulated by specific interactions with other *cis*-acting regulatory elements, such as the promoters of each of the two flanking genes, *Dlx5* and *Dlx6*. An overall distinct set of transcriptional activators binding to upstream and intergenic regulatory sequences would be responsible for the differences in *Dlx5* and *Dlx6* expression patterns. Experiments in zebrafish designed to examine the interactions between the intergenic forebrain enhancer and the *dlx4* and *dlx6* promoters are presently under way to address this issue. In summary, although it is possible that the

intergenic forebrain enhancer is shared between *Dlx5* and *Dlx6*, which would explain their partially overlapping patterns of expression, additional mechanisms must account for the overall distinct expression of the two genes in the forebrain.

Transgenic animals carrying constructs containing both Id4/d6i and Id4/d6ii always exhibit expression of the reporter gene in the ventral forebrain. Reporter expression is often seen in the olfactory placodes, but never in regions of the embryo, such as the branchial arches, the inner ear and the AER of the limb buds where *Dlx5* and *Dlx6* or their zebrafish orthologs are also expressed. This suggests, that elements necessary for proper expression in the latter areas are located outside the *Dlx5/Dlx6* intergenic region. On the other hand, several transgenic animals with the Id4/d6i or ID5/D6i constructs showed expression in the ectomesenchyme of the branchial arches reminiscent of endogenous *Dlx* expression. The mechanisms that underlie such results are, at present, unclear but may involve integration effects. It is possible that intergenic sequences, outside Id4/d6i (ID5/D6i) are necessary to restrict the activity of this enhancer to the ventral forebrain.

Dlx proteins interact with intergenic regulatory sequences

Expression of *Dlx5* and *Dlx6* is affected in the ventral forebrain of *Dlx1/2* null mutants (Anderson et al., 1997a,b). Thus, *Dlx5* and *Dlx6* transcripts are not detectable in the SVZ of the LGE at E12.5 and E14.5. The normal expression of the truncated *Dlx1* and *Dlx2* transcripts in forebrain cells of the mutant (Fig. 3.6 and data not shown)

suggests that the SVZ cells which normally express *Dlx5* and *Dlx6* are still present. Therefore, the reduced expression of *Dlx5* and *Dlx6* strongly suggests that *Dlx1* and/or *Dlx2* are required for the induction and/or maintenance of *Dlx5* and *Dlx6* expression. This might be achieved by direct regulation of *Dlx5/Dlx6* expression by the Dlx1 or Dlx2 proteins which can function as transcriptional activators (Liu et al., 1997; Yu et al., 1999; Zhang et al., 1997). Such cross-regulatory interactions involving homeobox genes have been described, for example for members of the *Hox* clusters (Gould et al., 1997; Nonchev et al., 1997; Studer et al., 1998) and for the zebrafish *dlx* genes (Zerucha et al., 1997). Alternatively, Dlx1 or Dlx2 may activate a yet unknown factor that is an essential regulator of *Dlx5* and *Dlx6* expression. The above two mechanisms are not mutually exclusive.

The zebrafish *Id4/d6* and *Id4/d6i* and the mouse *ID5/D6i* intergenic transgenes are expressed with patterns similar to the endogenous mouse *Dlx5* gene (Fig. 3.3, 3.5). Like endogenous *Dlx5* and *Dlx6*, the activity of the zebrafish *dlx4/dlx6* intergenic transgene (*Id4/d6*) is drastically reduced in the *Dlx1/2* null mutants (Fig. 3.6). These results suggest that the Id4/d6i and the orthologous ID5/D6i enhancer may be the site that is directly regulated by Dlx1 and Dlx2 *in vivo* in zebrafish and mouse, respectively. Our finding that the Dlx2 (Fig. 3.7A) and Dlx1 (not shown) proteins, of either zebrafish or mouse origin are able to up-regulate transcription of reporter constructs containing the conserved Id4/d6 intergenic sequences in co-transfection experiments supports the view that these sequences are the site of cross-regulatory interactions *in vivo*. Furthermore, we were able to demonstrate binding of Dlx2 to a short fragment of Id4/d6i in gel mobility shift assays.

This same fragment of Id4/d6i is responsible for most if not all of the activation by Dlx proteins in transfection experiments and is able to target transgene expression to the ventral forebrain in mice and zebrafish.

The fact that Dlx proteins are able to activate transcription of Id4/d6i-containing constructs in cultured cells does not mean that such interactions are sufficient for maximal enhancer activity *in vivo*. We cannot rule out that additional proteins endogenous to the transfected cells cooperate with Dlx in activating transcription of the co-transfected reporter constructs.

Most, if not all Dlx proteins are able to activate transcription of reporter constructs containing the Id4/d6i sequence similarly in co-transfection experiments (data not shown). The high degree of similarity in the DNA-binding homeodomains of the *Dlx* gene family makes it likely that this result indicates the ability of all Dlx proteins to bind this element. It is therefore difficult using co-transfection to determine which Dlx proteins bind to ID5/D6 sequences *in vivo*. It is quite possible that more than one of the four Dlx proteins present in forebrain cells interact with ID5/D6 (Id4/d6) in the normal embryo. In addition to cross-regulatory interactions with Dlx1 and/or Dlx2, it is also possible that the ID5/D6i sequences might be the site of auto-regulatory mechanisms involving Dlx5 and/or Dlx6. Such auto-regulatory loops have been described for other developmental genes encoding transcriptional regulators, including homeobox-gene products (Pöpperl and Featherstone, 1992). Furthermore, because of structural differences outside the homeodomain between Dlx family members, we cannot rule out

that different Dlx proteins might participate in specific protein-protein interactions *in vivo* and that these might influence their ability to participate in auto- or cross-regulatory interactions at the ID5/D6 sequences.

The Id4/d6 (ID5/D6) sequences have many potential homeodomain binding sites (Fig. 3.2). We consider it likely, for several reasons, that, in addition to the Dlx proteins, other homeodomain proteins and other types of transcription factors, may also modulate the activity of the Id4/d6 (ID5/D6) enhancer. First, we hypothesize that there could be repressor molecules that act on Id4/d6 to restrict its expression primarily to the forebrain and olfactory placode; in the absence of these repressors, the enhancer would be expressed in all cells containing the Dlx proteins. Almost all of the transgenic mice with Id4/d6 only have forebrain expression; whereas when smaller fragments are used (which may have eliminated the repressor binding sites), then additional sites of expression are revealed (e.g. branchial arches and limb AER; Fig. 3.5). Most of these additional sites correspond to tissues expressing Dlx proteins.

Our second reason for suggesting that the Id4/d6 enhancer is regulated by additional transcription factors is based on the high degree of sequence conservation (83%) between Id4/d6i (~ 400 bp) and orthologous enhancers in mouse and man (Fig. 3.2A). Based on the nucleotide sequence of this enhancer, we have not yet been able to unequivocally identify consensus sequences for distinct classes of transcription factors other than homeodomain proteins, although similarities with some binding sites do exist.

The biochemical basis for the partial functional redundancy observed between *Dlx1* and *Dlx2* (Anderson et al., 1997a; Qiu et al., 1997) can be understood in part based upon the similar transcriptional regulatory properties of the Dlx proteins with respect to the Id4/d6 and ID5/D6 enhancers in tissue culture cells (Fig. 3.7A and Yu et al., 1999). However, these results do not explain why the loss of *Dlx1* and *Dlx2* expression only eliminates Dlx5, Dlx6 and Id4/d6 expression in the SVZ, whereas their expression is maintained in the early-born mantle cells (Fig. 3.6). This observation indicates that other transcription factors (which might correspond to unknown Dlx proteins), regulate Dlx5, Dlx6 and Id4/d6 expression in a subset of early forebrain cells. Furthermore, the loss of Dlx5 and Dlx6 expression in the SVZ of the *Dlx1/Dlx2* null mutants also raises the possibility that the mutant phenotype may be attributable to the loss of function of all four genes, implying some functional redundancy between them. A better understanding of any differences in biochemical activities of Dlx proteins, such as involvement in specific protein-protein interactions, would help elucidate the functional consequences of the partially overlapping expression of these genes during development.

Intergenic region and *Dlx* gene evolution

The high degree of sequence similarity that we observed in the intergenic region between a pair of Dlx genes of mouse, of human, and of teleost fish (Fig. 3.2) is remarkable considering that these sequences are outside the coding regions of either genes. High degrees of sequence similarity outside gene coding regions have been observed previously between human and mouse sequences (see for example Beckers et

al., 1996; Williams et al., 1998) and some sequence conservation has also been found with sequences of distantly related vertebrates such as the pufferfish, *Fugu rubripes* and zebrafish (Beckers et al., 1996; Kimura et al., 1997; Marshall et al., 1994; Morrison et al., 1995). However, this is the first time such extended sequence similarity is found between a mammalian sequence and a teleost fish sequence outside the protein-coding regions of genes.

The inverted convergent configuration of pairs of *distal-less*-related genes is ancient as it has been reported for the ascidian *Ciona intestinalis* (Di Gregorio et al., 1995). The distance that separates the two genes is relatively small (2-10 kb) for all cases reported thus far. It is likely that the paired organization arose after the divergence of arthropods from the lineage that would give rise to vertebrates since insects are thought to have only one *distal-less* gene. It is possible that one or a few regulatory sequences found downstream of the *distal-less* gene in the common ancestor to modern day invertebrates and vertebrates were preserved after the first duplication and inversion event that produced the first pair of *distal-less/Dlx* genes. Enhancer sequences have been described downstream of the *Drosophila distal-less* gene (O'Hara et al., 1993; Vachon et al., 1992) and it will be interesting to determine if there is any degree of functional conservation in these enhancers and those described in the current study. A potential evolutionary advantage of enhancer-sharing by the two linked genes would be consistent with the conservation of the inverted convergent configuration and the maintenance of a relatively short intergenic distance. Enhancer sharing has been previously demonstrated

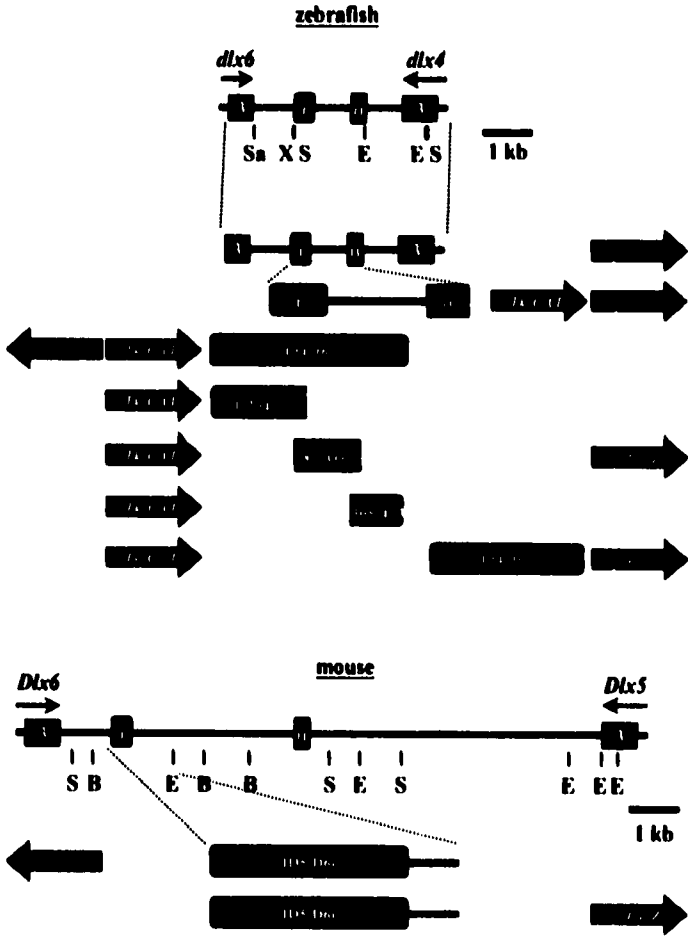
for some of the clustered *Hox* genes (Gould et al., 1997; Sharpe et al., 1998; van der Hoeven et al., 1996).

Vertebrates have at least three pairs of linked *Dlx* genes (Ellies et al., 1997b; Liu et al., 1997; McGuinness et al., 1996; Simeone et al., 1994), the *Dlx1/Dlx2*, *Dlx5/Dlx6* (*dlx4/dlx6* in zebrafish), and *Dlx3/Dlx7* pairs. The presence of conserved regulatory sequences may not be unique to the *Dlx5/Dlx6* gene pair as the *Dlx1/Dlx2* intergenic region also contains highly conserved sequences (Zerucha, Qiu, Liu, Rubenstein and Ekker, unpublished) although the roles of such sequences in *Dlx* gene regulation is, at present, unclear. The function and evolution of intergenic enhancer sequences, combined with studies of the functional specificity of *Dlx* genes, will enable us to understand the mechanistic basis for the concerted action of Dlx proteins in embryonic cells and the position of the *Dlx* genes in regulatory cascades during development.

ACKNOWLEDGMENTS

We thank Genny Giroux and Wei Lin for technical assistance, Allison Lewis for help in sequencing the zebrafish *dlx4/dlx6* intergenic region, Marie-Andrée Akimenko and Lucie Jeannotte for critical reading of the manuscript. This work was supported by grants from the Medical Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada and the March of Dimes Birth Defects foundation to M.E.

Figure 3.1. Genomic organization of the zebrafish *dlx4* and *dlx6* genes (top) and of the orthologous murine *Dlx5* and *Dlx6* genes (bottom) indicating the location of conserved sequences with putative regulatory function. The third exons of zebrafish *dlx4* and *dlx6* and of mouse *Dlx5* and *Dlx6* are represented by boxes. Direction of transcription is indicated by arrows. B, *Bam*HI; E, *Eco*RI; X, *Xho*I; S, *Sac*I; Sa, *Sal*I. The constructs for the production of transgenic animals and for transfection experiments are schematized. The position and orientation of the intergenic fragments relative to the reporter genes (*lacZ*, *CAT* and *GFP*) is shown. β, minimal β-globin promoter; tk, thymidine kinase promoter. Numbers of primary transgenic embryos or embryos from transgenic lines that show *LacZ* expression in various sites of *Dlx* expression are indicated to the right of each construct.



	TRANSGENIC MICE		TRANSGENE EXPRESSION			
	Primary	Lines	Ventral Forebrain	Olfactory Placode	Visceral Arches	Apical Ectodermal Ridge
	N.D.	7	N.D., 7	N.D., 5	N.D., 0	N.D., 0
	12	N.D.	12, N.D.	0, N.D.	3, N.D.	0, N.D.
	5	N.D.	5, N.D.	0, N.D.	0, N.D.	1, N.D.
	10	2	1, 0	1, 1	0, 2	1, 0
	4	4	4, 3	1, 1	2, 2	1, 0
	9	N.D.	9, N.D.	1, N.D.	5, N.D.	0, N.D.

Figure 3.2. Conserved sequences in the intergenic region that separates a pair of vertebrate *Dlx* genes. (A) Alignment of ID5/D6i sequences from human (h) and mouse (m) and Id4/d6i from zebrafish (z). (B) Alignment of ID5/D6ii sequences from human and mouse and Id4/d6ii from zebrafish. The human sequences were retrieved from GenBank; accession number AC004774. Complete sequence identity across the 3 species is indicated by a star. The nucleotide positions of Id4/d6i corresponding to the 187 - 316 fragment are shaded.

A

```
             26
             |
zfId4/d6i    GACAAAATGTTTT----TCCCCTTTCTCTCTGCCTTGATCTGC-AAGACGCTGAATGCCT
mID5/D6i     GA-AAAATGTTTTCTTTTCTTTTTTCCCCCTTCCT--A-CTGTGAA-ACTTTGGGTTCGT
hID5/D6i     GA-AAAATGTTTTCTTTTCTTTTTTCCCCCTTCCT--A-CTGTGAA-ACTTTGGGTTCGT
             ** **********    **    **** * ** ***  * ***  ** **  **  * * *

             81
             |
zfId4/d6i    TTCGCCACAGGATCGATCCTGAACAAAGCATCCAGCTGCAGT-CTCGTTCAATATGAAGG
mID5/D6i     AGCTCCCCAGGATCAATTCTGAACAAAGCCTCCAGCTGCAGTGC-CATCCAATTTGAAGC
hID5/D6i     AGCTCCCCAGGATCAATTCTGAACAAAGCCTCCAGCTGCAGTGC-CATCCAATTTGAAGC
               * ** ******* ** *********** *********** * * * **** *****

             140                                                187
             |                                                  |
zfId4/d6i    AGATATTTGGGACAATTTATGGGTTTTATCCA-AGAGAGGGTTTTTTTCCATTCTCATAA
mID5/D6i     AGACATTGGGGACAATTTAAGGTTTTTATCCACA-AGAAGGTTTTTTTCCATTCTCTTAA
hID5/D6i     AGACATTGGGGACAATTTAAGGTTTTTATCCACA-AGAAGGTTTTTTTCCATTCTCTTAA
             *** *** ********** ** ********* * *** ****************** ***

             199  204
             |    |
zfId4/d6i    ATGCAGACATAATTAGGGTAATTTTTGATGTAGCCCGCTGATTACAGCGTTTTTACCGTC
mID5/D6i     ATGCAGCCATAATTAGAGTAATTTTTCATGTAGCCCGCTGATTACAGCGTTTTTACCGTC
hID5/D6i     ATGCAGCCATAATTAGAGTAATTTTTCATGTAGCCCGCTGATTACAGCGTTTTTACCGTC
             ****** ********* ********* ********************************

             259                                        304         316
             |                                          |           |
zfId4/d6i    AAAGATAATTACCTGTAATTTTCTACCACTTTTAATACTAAAA-GGCATCTTTATTTGGA
mID5/D6i     AAAGATAATTACCTGTAATTTT---CCACTTTTAATACTAAAAAGCCATCTTTATTTAGA
hID5/D6i     AAAGATAATTACCTGTAATTTTCTTCCACTTTTAATACTAAAAAGCCATCTTTATTTAGA
             *********************    ***************** * *********** **

             318
             |
zfId4/d6i    TTT--GAAGTGGCAAAGACGAAACAAAAGATG-AAATTATTCGGTTATTCATGTACAAAT
mID5/D6i     TTCAGGAACAGG-AAAGGCGAAACAAAAGAGGGAAATTATTCTGTTATTCATACACAAAT
hID5/D6i     TTCAGGAACAGG-AAAGGCGAAAGAAAAGAGGGAAATTATTCTGTTATTCATACACAAAT
             **   ***  ** **** ***** ****** * ********* *********  ******

             375                                             420
             |                                               |
zfId4/d6i    TGCCGAGGA-G-AGTGAACGC-----TTGGGATATTATCTAAAATTACAATGT
mID5/D6i     TGCAGAG-ACGTAG-G-AC-CTAAAATTGAAA-ATTAACCAAAATTATAATGC
hID5/D6i     TGCAGAG-ACGTAG-G-AC-CTAAAATTGAAA-ATTAACCAAAATTATAATGC
             *** *** * * ** * ** *      ***  * **** * ******* ****
```

B

```
             1128
             |
zfId4/d6ii   GTAAGCAAATCCAGGTGTGAAAGTGTTTGCACACCCC-GACACCTCTTATATATTGCCTG
mID5/D6ii    GTGAGCACATCCAGGTGTGAAATTGTTTGCACACCCCAG-CACCTCTTATAT--TGCCAG
hID5/D6ii    GTGAGCACATCCAGGTGTGAAATTGTTTGCACACCCCAG-CACCTCTTATAT--TGCCAG
             ** **** ************** ************** * ***********   **** *

             1187
             |
zfId4/d6ii   CAAAATTAGCCGTTATTACTGTCACTGTTTAGTGATGG-------TGAGCCAGGG-----
mID5/D6ii    CAAAATTAGCTGTTATTACTGTCACTGTTTAGTGATGGTT-AGCGTGATACAAAAAAAAA
hID5/D6ii    CAAAATTAGCTGTTATTACTGTCACTGTTTAGTGATGGTTCAG-GTGGTACAAAAAAAAA
             ********** *************************        **   **

               1235
               |
zfId4/d6ii   ----AAAAAAC---------CTTCTGCAATCAAGAACCAGGCGCATCTTTGCAAATTATA
mID5/D6ii    AAAAAAAAAAAAAAAAAAAAGCTGCTGTAATCAAGA-CCTGGCGCATCTTTGCAAATTACA
hID5/D6ii    AAAAAAAAAAAAAAAAA---CTGCTGTAATCAAGA-CCTGGCGCATCTTTGCAAATTACA
                ******         ** *** ******** ** ****************** *

             1282
             |
zfId4/d6ii   GATAATTGTAAATGTCCAGATTATGATAAT-GGAG-CTAATCCTGG-TGGAAGTATAATT
mID5/D6ii    GATAATTGTAAACGTCCAGATTATGATAATAGCATCCTAATCCAGCCTGCAA-TATAATT
hID5/D6ii    GATAATTGTAAACGTCCAGATTATGATAATAGCATCCTAATCCAGCCTGCAA-TATAATT
             *********** **************** * *  ******* *  ** ** *******

             1340
             |
zfId4/d6ii   ATTGTTGAGTGTTACAGT-TGAGGGTGTCCAGTAAAGCTAACT--GTCATTATTTATACG
mID5/D6ii    AT-ACAGAGTGTTACA-TCTGAAACTGTCCAGTAGGGCTAATTCAGCCATTATTTAGACC
hID5/D6ii    ATTACAGAGTGTTACA-TCTGAAACTGTCCAGTAGGGCTAATTCAGCCATTATTTAGACC
             **    ********** * ***   *********  ***** *  * ********* **

             1396
             |
zfId4/d6ii   CGGTTT
mID5/D6ii    CTATTT
hID5/D6ii    CTATTT
             *  ***
```

Figure 3.3. A DNA fragment encompassing the zebrafish *dlx4/dlx6* intergenic region directs expression of *lacZ* in transgenic mouse embryos with patterns that closely recapitulate endogenous *Dlx5* and *Dlx6* expression in the forebrain. (A-C) *lacZ* expression in the ventral thalamus and telencephalon, and in the olfactory placodes in (A) E10, (B) E11, and (C) E12 whole mount mouse embryos. D) Coronal section of an E14.5 stage mouse embryo with *LacZ* expression in the lateral ganglionic eminence (LGE). Higher β-galactosidase activity is seen in the subventricular zone (SVZ) compared to the mantle (MZ). (E-F) *in situ* hybridizations with (E) *Dlx5* and (F) *Dlx6* probes on coronal sections adjacent to that seen in (D). Note that the relative patterns of *Dlx5* expression in the SVZ and MZ more closely resemble that seen in transgenic animals (D) than do the relative patterns of *Dlx6* expression. (G-I) More caudal sections of same stage embryos. Expression of β-galactosidase in the caudal ganglionic eminence (CGE) and in the preoptic area (POA). (H-I) in situ hybridizations with *Dlx5* and *Dlx6* probes, respectively on sections adjacent to those seen in G. Embryos in (A-B) are from line 7679 and those in (C-I) are from line 1469. (J-L) A 1.4 kb *Xho*I/*Eco*RI fragment from the zebrafish *dlx4/dlx6* intergenic region directs expression of a transgene which recapitulates endogenous *dlx* expression in the forebrain of zebrafish embryos. (J) Expression of Green Fluorescent Protein (GFP) directed by the 1.4 kb Id4/d6 zebrafish fragment in a 36 h embryo. The patterns are similar to the expression of the endogenous *dlx4* (K) and *dlx6* (L) genes. The domains I and II of *dlx* expression correspond, by analogy, to the diencephalic (I) and telencephalic (II) domains of *Dlx* expression in the mouse. Other abbreviations: I, domain I; II, domain II; AEP, anterior entopeduncular area; BT, basal telencephalon; Cx, cortex, HY, hypothalamus; LV, lateral ventricle; OB, prospective olfactory bulb; OP, olfactory placodes; Se, septum; SPV, supraoptic paraventricular area; VT, ventral thalamus; VZ, ventricular zone.

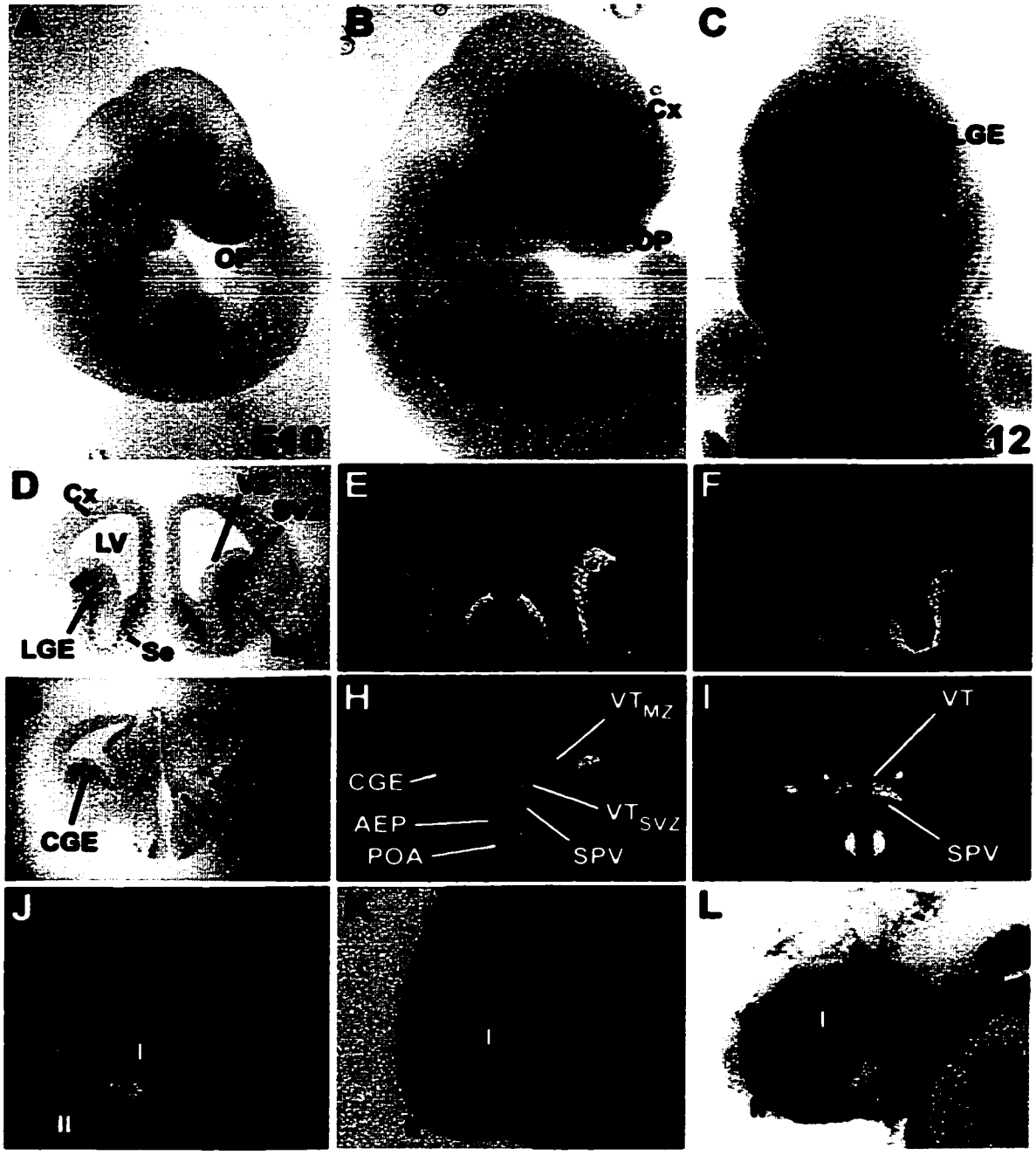

A
B
C
Cx
OP
LGE
12
D
Cx
LV
LGE
Se
E
F
CGE
H
VT_MZ
CGE
AEP
POA
VT_SVZ
SPV
I
VT
SPV
J
I
II
I
L
I

Figure 3.4. The *dlx1* and *dlx2* genes are expressed in more immature cells of the zebrafish forebrain than their *dlx4* and *dlx6* paralogs. Transverse sections of 48 h-old zebrafish embryos at the level of the telencephalon (A, C, E, G) and of the diencephalon (B, D, F, H) are shown with dorsal to top. Cells that express *dlx1* and *dlx2* are closer to the midline compared to those expressing *dlx4* or *dlx6*. The more medial expression of *dlx2* compared to *dlx4* confirms our previous observation (Akimenko et al., 1994).

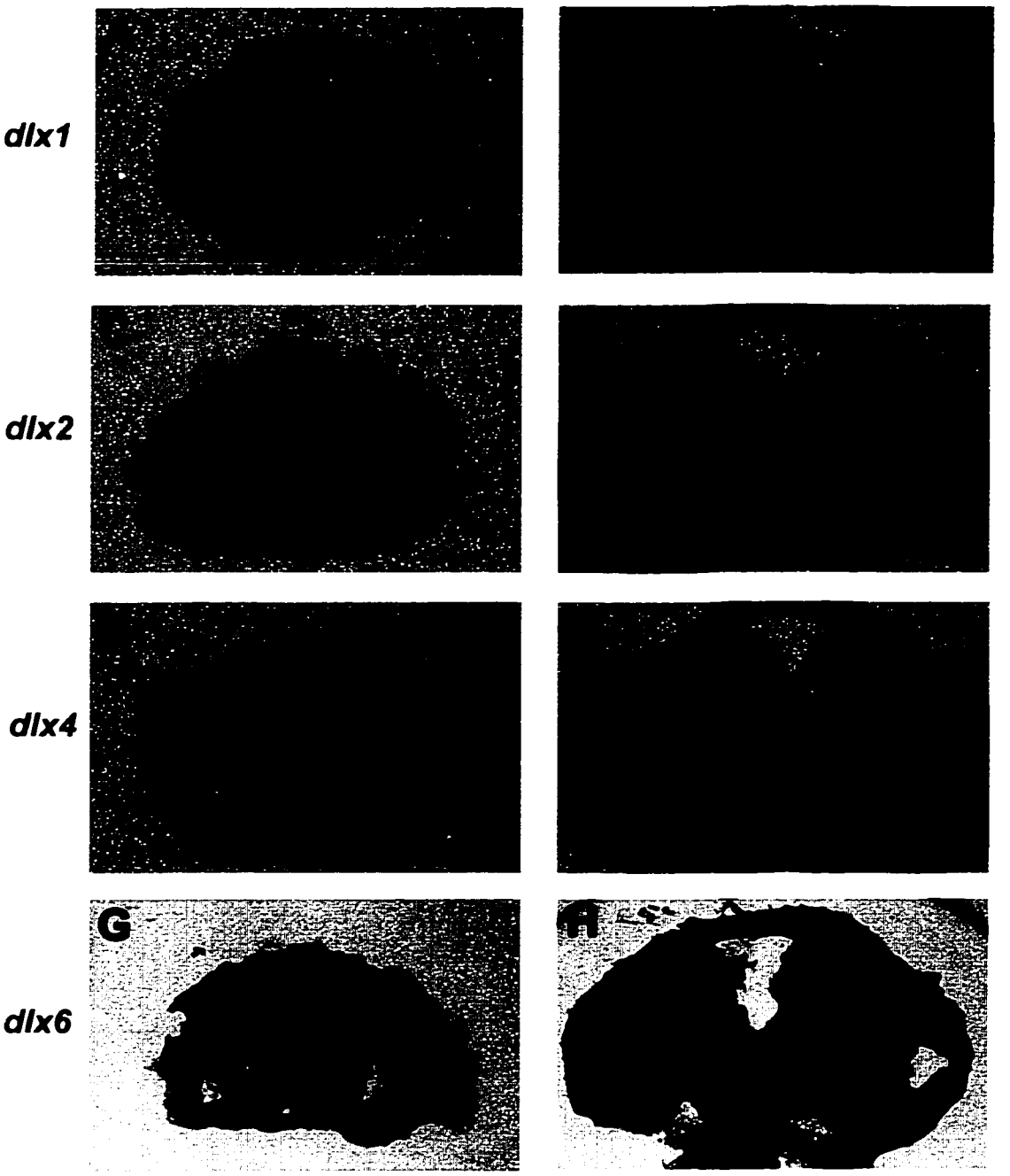

dlx1
dlx2
dlx4
dlx6
G
A

Figure 3.5. Specific intergenic sequences from either mouse or zebrafish target gene expression to the forebrain of E11 mouse embryos with highly similar patterns. (A) zebrafish Id4/d6i . (B) Mouse ID5/D6i. In addition to the forebrain, β-galactosidase activity was observed in the first two branchial arches in 2 of 3 lines and 2 of 4 primary transgenic embryos. (C) 187-316 fragment of zebrafish Id4/d6i. In addition to the forebrain, β-galactosidase is also expressed in the apical ectodermal ridge (AER) of the limb buds in 1of 5 primary transgenic embryos carrying this construct. Note that the forebrain expression patterns in (A-C) are highly similar to those of Fig 3.3B (full zebrafish intergenic fragment). Abbreviations: Mx, maxillary component of the first branchial arch; Md, mandibular component of the first branchial arch; Hy, hyoid arch. Other abbreviations as in Fig. 3.3.

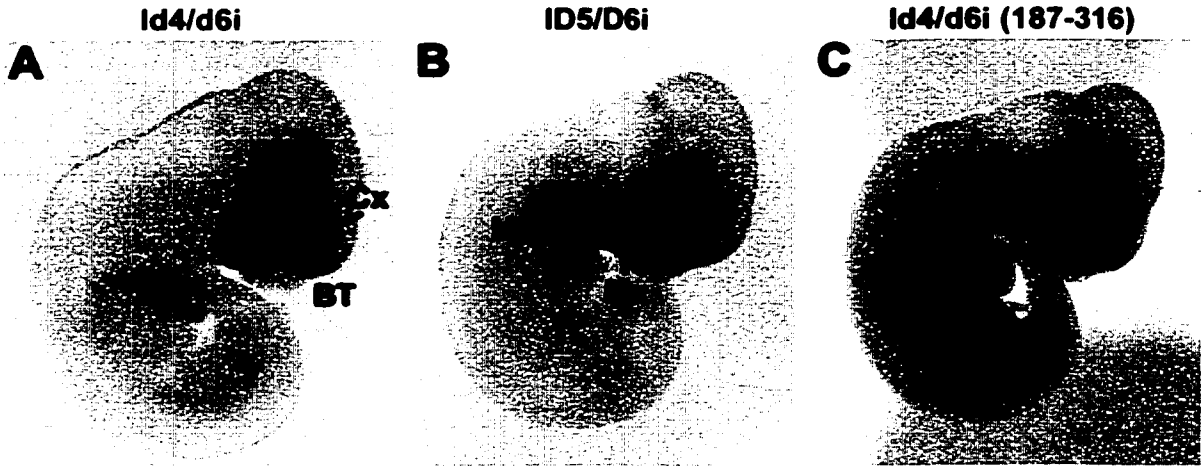

Id4/d6i
A
Cx
BT
ID5/D6i
B
Id4/d6i (187-316)
C

Figure 3.6. Forebrain expression of a transgene driven by Id4/d6 is drastically reduced in mice with a targeted null mutation of the *Dlx1* and *Dlx2* genes. (A) Coronal section showing *lacZ* expression (arrow) driven by Id4/d6 in the telencephalon of a E12.5 *wt* mouse embryo. (B) Section adjacent to that in (A) from a mouse embryo with the Id4/d6 transgene and a null mutation for *Dlx1* and *Dlx2*. (C) *LacZ* expression driven by Id4/d6 in the telencephalon (arrow) in wt E 14.5 mouse. (D) reduced *LacZ* expression in an adjacent coronal section to that shown in (C) in a mouse with a null mutation in the *Dlx1* and *Dlx2* genes. (E-H) *in situ* hybridizations with a *Dlx5* probe on coronal sections roughly adjacent to those shown in (A-D). (I-L) *in situ* hybridizations with a *Dlx6* probe. (M-P) *in situ* hybridizations with a full-length probe which recognizes full length *Dlx1* as well as a truncated form of *Dlx1* expressed in the *Dlx1/Dlx2* null mutant mice. The asterisk indicates that *Dlx5*, *Dlx6* and *zfId4/Id6lacZ* expression is maintained in the POA/SE in the *Dlx1/Dlx2* mutants. Abbreviation: MGE, medial ganglionic eminence. Other abbreviations as in Fig. 3.3.

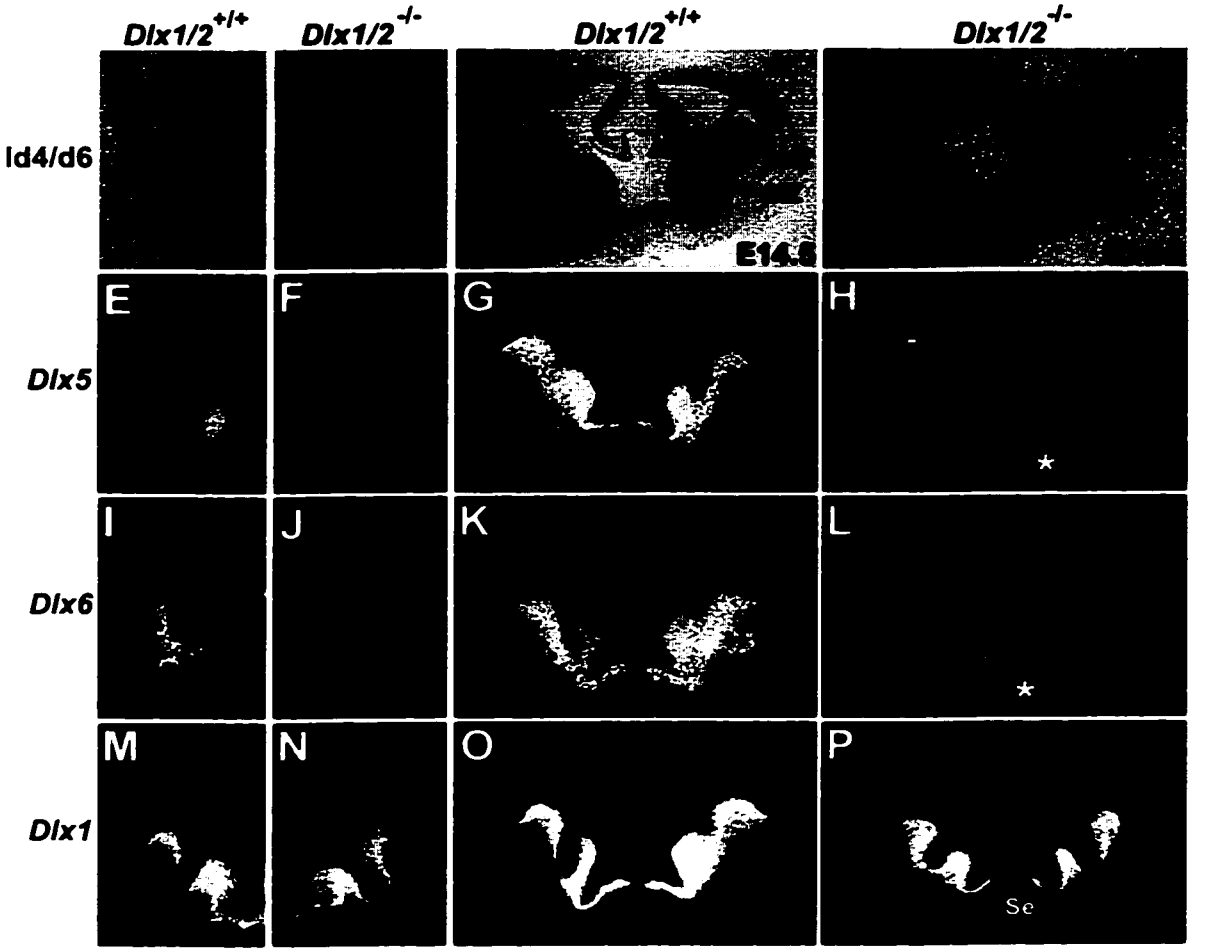
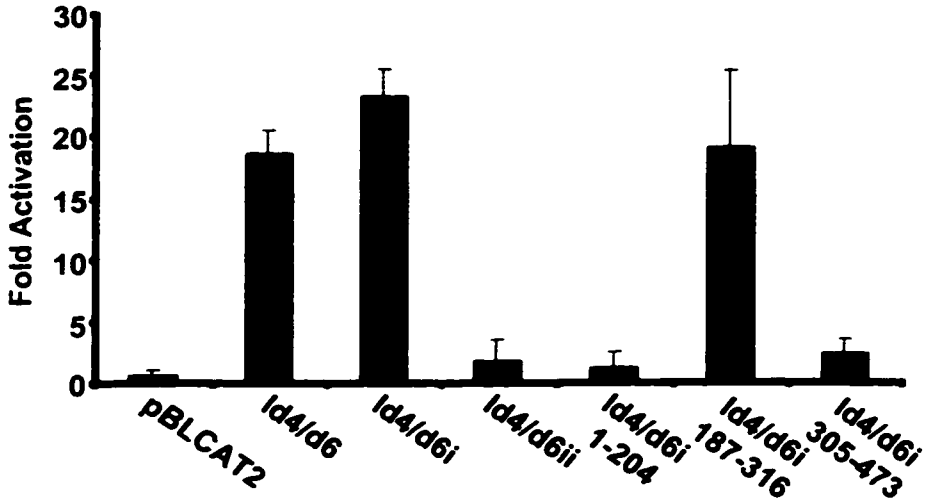
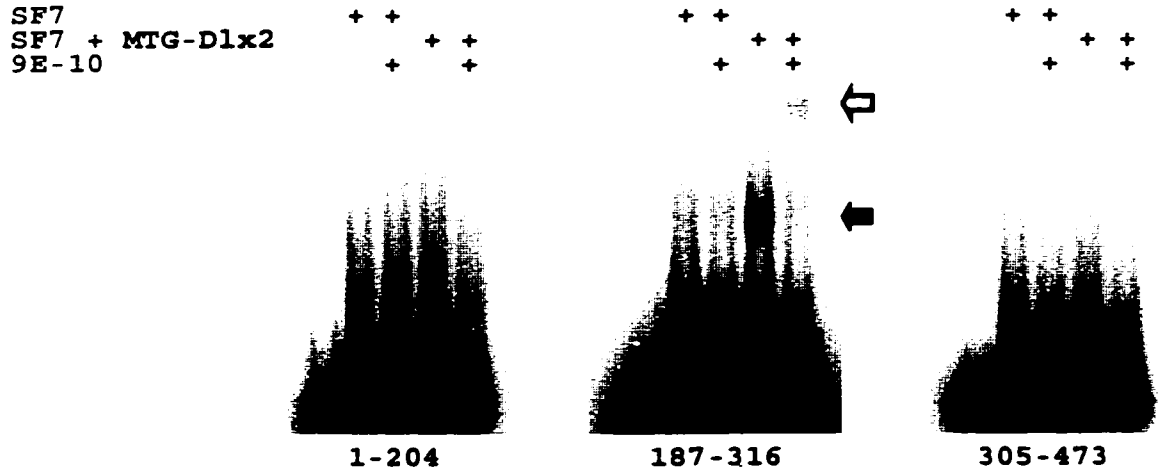

Dlx1/2+/+
Dlx1/2-/-
Dlx1/2+/+
Dlx1/2-/-
Id4/d6
E14.5
E
F
G
H
Dlx5
I
J
K
L
Dlx6
M
N
O
P
Dlx1
Se

Figure 3.7. (A) The zebrafish Dlx2 protein can activate transcription through intergenic regulatory sequences in transient transfection assays. Co-transfected Dlx2 activates transcription through the 1.4 kb *Xho*I/*Eco*RI fragment from the zebrafish *dlx4/dlx6* intergenic region (Id4/d6), through Id4/d6i, but not Id4/d6ii. The 187-316 fragment from Id4/d6i (see Fig. 3.2A) is also a target for Dlx2 but neither are the 1-204 nor the 305-450 fragments. All values shown represent fold activation in the presence of Dlx2 relative to the same construct in the absence of co-transfected Dlx2. The P19 embryonic carcinoma (EC) cell line was transfected by the calcium-phosphate precipitation procedure. The values shown represent three independent experiments ± standard error on the mean. (B) The zebrafish Dlx2 protein binds the 187-316 fragment of Id4/d6i in a gel mobility shift assay. Three fragments from Id4/d6i: 1-204; 187-316, 305-450 were radiolabeled and incubated with a nuclear extract from a SF7-derived cell line which expresses MTG-Dlx2 or with a control SF7 nuclear extract. A lower mobility complex is indicated by the solid arrow. In the presence of the 9E-10 antibody directed against the MTG epitope of MTG-Dlx2, this complex is supershifted (open arrow).

A


Fold Activation
30
25
20
15
10
5
0
pBLCAT2
Id4/d6
Id4/d6i
Id4/d6ii
Id4/d6i 1-204
Id4/d6i 187-316
Id4/d6i 305-473

B


SF7
SF7 + MTG-D1x2
9E-10
1-204
187-316
305-473

4. FUNCTIONAL CONSERVATION OF A SERIES OF *DISTAL-LESS* ENHANCER ELEMENTS DURING VERTEBRATE AND INVERTEBRATE EVOLUTION

Unpublished

ABSTRACT

The Vertebrate *Dlx* gene family consists of 3 linked pairs of convergently transcribed genes and in zebrafish at least two additional apparently unlinked genes. The linked organization of *Dlx* genes is ancient as it is also observed in urochordates. The linked *Dlx* genes demonstrate a striking degree of overlap in expression patterns during development. Previously we have described elements in the intergenic region of one of these gene pairs that is well conserved between teleost fish and mammals and that may at least partially account for the overlap in expression patterns of its neighboring *Dlx* genes by acting as a shared enhancer element. Here we describe similar elements in the intergenic regions of the other linked gene pairs. These elements are able to specifically direct expression in transgenic mice and remarkably at least half of them can also direct specific expression in analogous regions in transgenic flies. Interestingly the elements which exhibit some function in transgenic flies are also the sites of regulation by the Dlx proteins themselves. Our results suggest a mechanism by which duplicated genes may be stably maintained during evolution as well as a possible explanation for the maintenance of the *Dlx* genes as linked pairs in the genome.

INTRODUCTION

Many of the genes involved in vertebrate ontogeny are members of multi-gene families. This is consistent with Ohno's proposal (Ohno, 1970) that polyploidization events which occurred early in the history of the vertebrate lineage were directly involved in the morphological diversification of this group. Probably the most well-understood example of a gene family that reflects this is the *Hox* family. The *Hox* genes have undergone a number of gene duplication and cluster duplication events during their evolution which has resulted in four clusters of ~40 genes in mammals and seven clusters of at least 47 genes in teleost fish such as the zebrafish. It appears that the chromosomal duplications that gave rise to the Hox genes also gave rise to the linked *Dlx* family of homeobox-containing genes (Stock et al., 1996).

The *Dlx* gene family comprises six members in mammals and at least eight in the zebrafish. The six orthologous genes between mammals and zebrafish are organized as convergently transcribed pairs in the genome, *Dlx1*/*Dlx2*, *Dlx3*/*Dlx7* and *Dlx5* (zebrafish *dlx4*)/*Dlx6* (Fig. 4.1A). The two additional zebrafish *Dlx* genes, *dlx5* and *dlx8* are not linked to each other and it is presently unknown if they constitute additional pairs with as of yet unidentified *dlx* genes. The paired organization of *Dlx* genes arose early in the history of the vertebrate lineage as it has also been reported in urochordates such as *Ciona intestinalis* (Di Gregorio et al., 1995). In the cephalochordate amphioxus however, whose lineage is thought to have diverged from that of vertebrates following the urochordate divergenece, only one *Dlx* homologue has been reported (Holland et al.,

1996). Lineages that diverged before urochordates, such as arthropods and nematodes, are thought to have one *Distal-less* gene. The vertebrate *Dlx* genes have been further classified into one of two clades: one consisting of *Dlx1*, *Dlx6*, *Dlx7* and zebrafish *dlx8*; and the other *Dlx2*, *Dlx3*, *Dlx5*(zebrafish *dlx4*) and zebrafish *dlx5*. Each *Dlx* gene pair consists of one member from each clade. The evolution of the *Dlx* genes, then, has been proposed to have involved a tandem duplication event prior to the divergence of tunicates, which resulted in an ancestral convergently transcribed gene pair. A number of subsequent chromosomal duplications following the urochordate/cephalochordate divergence then gave rise to the modern *Dlx* family as observed in mammals and teleost fish.

The domains of expression patterns of the *Dlx* genes have been fairly well conserved amongst vertebrate species. Strikingly, the expression patterns of the linked Dlx genes have considerable overlap and include regions of the forebrain, olfactory placodes, visceral arches and limbs/fins. This suggests that many of the *cis*-regulatory elements and mechanisms that direct *Dlx* expression have been conserved at least throughout vertebrate evolution. Interestingly, expression of the *Drosophila Dll* has been reported in the developing legs and maxillary appendages that may be considered as loosely analogous to the visceral arch-derived jaw structures of vertebrates.

The conserved and divergent patterns of *Dlx* gene expression, across diverse phyla, suggests a mechanism which may have contributed to the evolution of this gene family. As genes are created by duplication events, for the new genes to be maintained,

there must be some selective pressure placed upon them. This issue has been examined in the past and was most recently discussed in terms of the maintenance of functional redundancy, a common by-product of gene duplications (Cooke et al., 1997; Nowak et al., 1997). Another possibility exists however, which may not be conducive to maintaining redundancy. This involves the coincident duplication of cis-regulatory elements as the coding regions of the genes are duplicated, followed by successive losses of regulatory elements from specific genes. Thus while the full or close to full complement of *cis*-regulatory elements remain associated with the gene family as a whole, each gene becomes associated with only a subset of these regulatory elements.

To address this, here we describe the identification of conserved vertebrate enhancer elements located in the intergenic regions of the linked *Dlx* genes. In transgenic mice, these elements are able to recapitulate specific domains of endogenous *Dlx* expression. Amazingly in transgenic flies, these same elements are able to recapitulate domains of expression of the single *Drosophila Dll* which correspond to potentially homologous regions observed in transgenic mice.

MATERIALS AND METHODS

Identification Of Conserved Sequences In The Zebrafish And Mouse *Dlx1/Dlx2* And *Dlx3/Dlx7* Intergenic Regions

Restriction fragments of a genomic clone containing the zebrafish *dlx1/dlx2* locus (Ellies et al., 1997b) were radiolabeled and hybridized to a Southern blot of various restriction digests of a mouse genomic clone containing the orthologous *Dlx1/Dlx2* locus (Liu et al., 1997). Of the zebrafish restriction fragments containing the *dlx1/dlx2* intergenic region, only a 440 bp *Sph*I-*Xho*I fragment hybridized to the mouse genomic fragments (Fig. 4.1A,B). This zebrafish fragment and the hybridizing mouse fragments were sequenced using the dideoxy procedure. The nucleotide sequence of the zebrafish *dlx3/dlx7* intergenic region was determined using the dideoxy procedure. and the corresponding mouse and human *Dlx3/Dlx7* intergenic region nucleotide sequences (Fig. 4.1A,C) were kindly provided by Frank Ruddle, Ken Weiss and David Stock. Nucleotide sequence comparisons were done using the CLUSTAL W version 1.7 (Thompson, 1984) multiple sequence alignment program.

Production Of Transgenic Mice

DNA fragments from the zebrafish or mouse *Dlx1/Dlx2* locus or the zebrafish *dlx3/dlx7* intergenic region were sub-cloned into the p1229 or p1230 *LacZ* reporter vectors (Yee and Rigby, 1993). For the production of transgenic mice, linearized

plasmids were injected at a concentration of 5 ng/μl in eggs from FVB/n crosses using standard procedures (Hogan et al., 1986). Transgenes were analyzed in either founder embryos or from established transgenic lines. Presence of the transgene was determined by PCR on DNA prepared from extra-embryonic tissues with the following oligonucleotide primers 5'-AGGGCAGAGCCATCTATTGC-3' and 5'-CGCTCATCCGCCACATATCC-3' derived, from the β-*globin* promoter and Lac-z sequences of the p1229/p1230 vectors respectively. As a positive control in these PCR reactions, a region of the mouse fetal hemoglobin gene was amplified with the oligonucleotide primers X1 (5'-GAT CAT GAC CGC CGT AGG-3') and X2 (5'-CAT GAA CTT GTC CCA GGC TT-3').

Morphological Analysis Of Transgenic Mice

Founder transgenic embryos or embryos from the cross of a transgenic male and a normal *FVB or CD1* female were harvested at various embryonic stages. Transgene expression was also determined in newborn pups, young mice and adults from an established line. Embryos were fixed in 1% formaldehyde; 0.2% glutaraldehyde; 0.02% NP-40 in phosphate-buffered saline (PBS) for 30 min at 4° C, washed in PBS for 20 min at room temperature and stained for β-galactosidase activity overnight at 28° C in a solution of 1 mg/ml X-gal, 5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6$, 2 mM $MgCl_2$, 0.02% NP-40 in PBS.

Production Of Transgenic *Drosophila*

The 440 bp *SphI/XhoI* Id1/d2 (Fig. 4.1A,B) and the 1.4 kb *XhoI/EcoRI* Id4/d6 (Fig. 4.1A and (Zerucha et al., 1999)) DNA fragments from the zebrafish *dlx1/dlx2* and *dlx4/dlx6* loci respectively, were sub-cloned into the pWHIZ vector (personal communication, S. Côté) containing an hsp70 basal promoter fused to the *LacZ* gene. For the production of transgenic *Drosophila*, plasmids were injected into dechorionated embryos (5% Sodium Hypochlorate) at a concentration of 600 ng/µl according to Spradling and Rubin (1982) except the source for transposase was 150 ng/µl pTurbo helper phage instead of the delta2-3 chromosome. Presence of the transgenes was confirmed by PCR amplification of the zebrafish Id1/d2 and Id4/d6 sequences and the *LacZ* gene. PCR was performed with oligonucleotides 245 (see below)/*LacZ* (5'-AATCGCCTTGCAGCACATCC-3') for Id1/d2 and 1060 (Zerucha et al., 1999)/*LacZ* for Id4/d6.

Morphological Analysis Of Transgenic *Drosophila*

Transgenic embryos were collected at various stages up to approximately stage 12 (7-9 hours) of development. Embryos were dechorionated in 6% sodium hypochlorate then fixed in 0.8% paraformaldehyde/PBS, 80% heptane. Embryos were washed once in: 10 mM Na2HPO4/NaH2PO4 pH 8.4; 150 mM NaCl; 1 mM MgCl2 and then again in the same solution with 0.2% Triton X-100. Following these washes, embryos were stained at 37°C, 16-24 h in: 10 mM Na2HPO4/NaH2PO4 pH 8.4, 150 mM NaCl, 1 mM MgCl2, 3

mM K3Fe(CN)6, 3 mM K4Fe(CN)6 and 0.2% 5-bromo-4-chloro-3-indolyl-β-D-galactopyranosid in N,N dimethylformamide.

Constructs For Transient Co-Transfection Experiments

The construction of the effector plasmid that expresses the zebrafish *dlx2* gene under control of the SV40 early promoter has been described previously (Zerucha et al., 1999). Reporter plasmids were constructed by inserting fragments of the zebrafish *dlx1/dl2* or *dlx3/dlx7* intergenic region into the pBLCAT2 vector (Luckow and Schütz, 1987) which contains the Thymidine Kinase (tk) minimal promoter driving expression of the Chloramphenicol Acetyl Transferase (CAT) gene. The 440 bp *Sph*I-*Xho*I Id1/d2 fragment from the zebrafish *dlx1/dlx2* intergenic region (Fig. 4.1A,B) was subcloned into pBLCAT2 directly upstream of the tk promoter. The 1.4 kb *Hind*III/*Sac*II fragment from the *dlx3/dlx7* intergenic region and which contains the 89 bp Id3/d7 element (Fig. 4.1A,C) was also subcloned into the pBLCAT2 plasmid. The deletions of Id1/d2 were prepared by PCR from a pSP72 (Promega) subclone, containing the 440 bp *Sph*I-*Xho*I Id1/d2 fragment using the following oligonucleotides:

133: 5'-CAGTTGAGCATTCTGGCT-3'; 245: 5'-TAGCTGGAAATTGTCTTG-3'; 323: 5'-GATGTTATGCTAAAA-3'; 328: 5'-TCCGCGATAATTATAT-3' and inserted in pBLCAT2 directly upstream of the tk promoter. The numeric name of the oligonucleotide refers to the position within Id1/d2 corresponding to the 5' end of the oligonucleotide (Fig. 4.1B) and the PCR generated deletion fragments were: T7/245, 133/SP6, 133/323, 328/SP6. The T7 oligonucleotide recognizes a region in pSP72

upstream of the Id1/d2 *Sph*I site which corresponds to position 1, and the SP6 oligonucleotide recognizes a region in pSP72 downstream of the Id1/d2 *Xho*I site which corresponds to position 440 (Fig. 4.1B).

Transient Co-Transfection Experiments

Transient co-transfection experiments were performed in the P19 murine embryonic carcinoma (EC) cell line as described previously (Zerucha et al., 1997; Zerucha et al., 1999). Cells were seeded 24 h prior to transfection at a density of 10^7 cells per 100 mm dish. Transfections were carried out by the calcium phosphate precipitation procedure (Sambrook et al., 1989). A total of 10 μg of DNA per dish was used in each transfection. This included: 2 μg pRSV-βgal as an internal control for transfection efficiency; 2 μg of reporter plasmid; 2 μg of effector plasmid; and sheared calf thymus DNA (Boehringer Mannheim) to the total of 10 μg. Precipitates were left on the cells for 16 h and the cells harvested 64 h post-transfection. Cells were collected in PBS, pelleted by centrifugation and resuspended in freeze/thaw buffer (250 mM Tris-HCl pH 8, 10 mM DTT, 15% Glycerol). Cell extracts were prepared by repeated cycles of freezing/thawing. β-Galactosidase activity was assayed as described by Sambrook et al. (1989). CAT activity was determined by thin-layer chromatography and measured as percent conversion of mono- and diacetylated chloramphenicol relative to unmodified plus acetylated chloramphenicol using the BIO-RAD GS-525 Molecular Imager System. CAT activity was standardized to β-Galactosidase levels to compensate for variations in transfection efficiency. Experiments were performed in duplicate and repeated a

minimum of three times. Error bars in the figures represent standard errors on the means of all replications.

Electrophoretic Mobility Shift Assays

A 132 bp DNA fragment corresponding to positions 328-440 of Id1/d2 (Fig. 4.1B) was amplified by PCR using the oligonucleotide primers 328 and SP6. The PCR product was cloned into the pCRII vector (Invitrogen). The fragment was excised with *Eco*R1 prior to filling of the 5' overhangs with the large fragment of DNA polymerase I (Klenow) in the presence of radiolabelled [α^{32}P]dATP. Binding reactions were carried out in a total volume of 20 µl in 12mM Hepes-KOH (pH 7.9), 1mM EDTA, 0.4 mM $MgCl_2$, 100 mM NaCl, 0.6 mM DTT, 0.6 mM PMSF, 13% glycerol. Nuclear extract (10 µg) from the stable SF7-MTG-Dlx2 cell line (Zerucha et al., 1999) or an equivalent prepared from control SF7 cells was pre-incubated with the 9E-10 anti-myc monoclonal antibody (Santa Cruz Biotech) or an equivalent volume of water at room temperature for 30 min. Following pre-incubation, 1 µg BSA, 1 µg sheared calf-thymus (Boehringer Mannheim) and 15000 cpm of radiolabelled probe were added and incubated at room temperature for 20 min. Protein-DNA complexes were resolved on a 4% polyacrylamide (29:1 acrylamide:bis-acrylamide) gel run in 1XTBE.

RESULTS

Identification Of Highly Conserved Elements In The Paired *Dlx* Intergenic Regions

Previously we have reported the presence of two highly conserved elements in the intergenic region of the *Dlx5* (zebrafish *dlx4*) and *dlx6* genes (Zerucha et al., 1999). These elements are conserved between zebrafish, mice and humans. To determine if similar elements exist in the intergenic regions of the other *Dlx* gene pairs, these regions (from between *Dlx1/Dlx2* and *Dlx3/Dlx7*) were compared between zebrafish and mouse by cross-species Southern hybridizations and nucleotide sequence alignments. A 440 bp region was identified in the *Dlx1/Dlx2* intergenic regions that is 91% identical between mouse and zebrafish and which we have named Id1/d2 and ID1/D2 for the zebrafish and murine elements respectively (Fig. 4.1A,B). A 117 bp region was identified in the *Dlx3/Dlx7* intergenic region by comparing the nucleotide sequence of this entire intergenic region in zebrafish, human and mouse (Webb Miller, personal communication). We have named this element Id3/d7 in zebrafish and ID3/D7 in the aforementioned mammalian species. The Id3/d7 element is 65% identical between zebrafish and mouse and 64% similar between zebrafish and human, the human and mouse sequences are identical apart from three single base pair substitutions (Fig. 4.1C).

The *Dlx* Intergenic Elements Are Able To Target Reporter Gene Expression In Transgenic Mice

We have previously demonstrated the Id4/d6-ID5/D6 elements to direct expression to the ventral forebrain of the developing mouse (Fig. 4.2B and Zerucha et al., 1999). To determine if the Id1/d2 and Id3/d7 elements are able to direct expression in a similar manner, we produced transgenic mice carrying the *LacZ* reporter gene coupled to either one of these elements. A construct containing the *LacZ* reporter gene under the control of a β-Globin minimal promoter and the zebrafish or murine Id1/d2-ID1/D2 element was injected into fertilized mouse eggs to produce transgenic mice. We obtained 4 lines and 1 primary transgenic animals carrying the murine ID1/D2 element and 1 line and 3 primaries carrying the zebrafish Id1/d2 element. Both of these orthologous elements direct virtually identical expression to the mandibular arch of transgenic mouse embryos (Fig. 4.2A and data not shown). This domain of expression within the visceral arches is consistent with the pattern observed for the endogenous *Dlx* genes although the transgene appears to be expressed in a slightly smaller region of the domain observed for the endogenous *Dlx* genes.

Similarly, we tested the ability of the full zebrafish *dlx3/dlx7* intergenic region to direct expression of the *LacZ* reporter gene under control of the same β-Globin minimal promoter. We obtained 2 lines and 2 primary transgenic animals. The entire *dlx3/dlx7* intergenic region directed expression to the genital ridge in transgenic mouse embryos

(Fig. 4.2C). We also tested the ability of the Id3/d7 element alone to direct expression of *LacZ* with no success to date.

The *Dlx* Intergenic Elements Are Able To Target Reporter Gene Expression In Transgenic *Drosophila*

It has been demonstrated previously that enhancers of the *Hox* genes have some activity in *Drosophila* (Gould et al., 1997; Haerry and Gehring, 1997; Haerry and Gehring, 1996; Keegan et al., 1997). To determine if the conserved vertebrate *Dlx* elements might echo this, we prepared lines of transgenic *Drososphila* carrying the *LacZ* reporter gene under the control of the hsp70 basal promoter and the conserved zebrafish intergenic elements. The Id4/d6 elements direct expression in transgenic *Drosophila* to the developing brain (Fig. 4.2E), what appears to be the peripheral nervous system (Fig. 4.2F) and the developing maxillary (not shown, but virtually indistinguishable from that shown in Fig. 4.2D). The Id1/d2 element directs expression of the reporter transgene to an anterior region of the embryo which will ultimately give rise to the maxillary (Fig. 4.2D). Remarkably the expression domains directed by the zebrafish *dlx* intergenic elements in *Drosophila* are very similar by analogy to the domains of expression they direct in transgenic mice: the Id4/d6 elements to the brain; and Id1/d2 to the developing jaw in mice and primordia for the mouth appendages in *Drosophila*.

Dlx Proteins Can Bind And Activate Transcription Through The Id1/d2 Element

Of particular interest is the ventral maxillary expression directed by Id1/d2 in *Drosophila*. Previously, a number of enhancer elements were reported for the *Dll* gene in *Drosophila* including one, ETD6, located downstream of the *Dll* coding region (O'Hara et al., 1993; Vachon et al., 1992) as are the vertebrate intergenic elements we have identified. In transgenic fly experiments, the ETD6 element directs expression to the ventral maxillary in a pattern virtually identical to that defined by Id1/d2 (O'Hara et al., 1993) suggesting that this element may have at least been functionally conserved throughout the evolution of vertebrates and arthropods. It has been determined that ETD6 is regulated by the *HOM-C* gene *Deformed* (*Dfd*). To determine if this may also be true for the Id1/d2 element, we tested the ability of one of the zebrafish Dfd orthologues, hoxd4, to regulate transcription through Id1/d2 in transient co-transfection experiments. It was determined that neither Hoxd4 alone, nor together with Pbx, a known partner which acts with the murine HoxD4 (Phelan and Featherstone, 1997), were able to activate transcription through Id1/d2 (not shown). Previously we have demonstrated that Dlx proteins are able to auto- and/or cross-regulate each other (Zerucha et al., 1997; Zerucha et al., 1999), including acting through the Id4/d6i element (Zerucha et al., 1999). To determine if Id1/d2 can be regulated by Dlx proteins as well, we performed similar co-transfection experiments and determined that all zebrafish *dlx* genes examined (*dlx1*, *dlx2*, *dlx3*, *dlx4*, *dlx6*) were able to activate transcription through the Id1/d2 element (Fig. 4.3 and data not shown). A number of zebrafish orthologues of the *Drosophila sine oculus* (*so*) gene *six3*, *six6*, and *six7* (kindly provided by Hee-Chan Seo;

Seo et al., 1998a,b) were also examined for the ability to regulate transcription through the Id1/d2 element, but to no avail.

To attempt to determine a specific region of Id1/d2 recognized by Dlx proteins to activate transcription, a number of deletion fragments of Id1/d2 were prepared (Fig. 4.3A) and co-transfected with a *dlx2* expression construct. Dlx2 is able to activate transcription through the 133-440 fragment but not the 1-245 fragment (Fig. 4.3B). Co-transfections with the further sub-divisions of the 133-440 fragment into the 133-323 and 328-440 fragments illustrated that Dlx2 is activating transcription through the 328-440 region of Id1/d2. In order to determine if Dlx proteins can directly interact with the 328-440 fragment of Id1/d2 as was suggested by their ability to regulate transcription through it, electrophoretic mobility shift assays (EMSAs) were performed. Nuclear extract from the SF7-MTG-Dlx2 cell line (Zerucha et al., 1999), which contains the zebrafish Dlx2 protein fused to six repeats of a c-myc epitope, was compared to nuclear extract from the SF7 parental cell line for the presence of proteins with the ability to bind the 328-440 fragment of Id1/d2. A retarded mobility complex was obtained with the SF7-MTG-Dlx2 nuclear extract (Fig. 4.4). This complex was further delayed in the presence of the 9E-10 monoclonal antibody which recognizes the MTG epitope of the MTG-Dlx2 fusion protein indicating that the lower mobility complex does indeed contain MTG-Dlx2.

DISCUSSION

Highly Conserved Elements From The Intergenic Regions Of The Paired Vertebrate *Dlx* Genes Can Target Expression In Transgenic Mice

Conserved elements have now been identified in the intergenic regions of all vertebrate, paired *Dlx* genes. These elements are well conserved between teleost fish and mammals. Of the four elements identified, at least two, Id1/d2 and Id4/d6i, are able to direct expression of a transgene in a pattern which represents a partial recapitulation of endogenous murine *Dlx* expression. Thus Id1/d2 and Id4/d6i direct expression to the visceral arches and ventral forebrain, respectively, of the developing mouse. It seems that the function of these two elements has been conserved during vertebrate evolution as the Id4/d6i element can function similarly in transgenic zebrafish (Zerucha et al., 1999) and preliminary evidence suggests this may also be true for the Id1/d2 element (Westerfield and Ekker unpublished observations). The roles of the Id4/d6ii and Id3/d7 elements are less clear. Previously we described the low activity of the Id4/d6ii element in transgenic mice, and while we have obtained transgenic mice with the entire *dlx3/dlx7* intergenic region, and thus which also contain the Id3/d7 element, we can not at this time confidently attribute the expression pattern conferred by the entire intergenic region to the Id3/d7 element. In comparing the full intergenic region of the murine and zebrafish *Dlx3/Dlx7* genes, the Id3/d7 element is the only conserved region, however this does not preclude the possibility that sequences other than the conserved elements may also be able to direct expression during development. To date, transgenic mice carrying just the

Id3/d7 element have exhibited no expression of the transgene. A possible explanation for this is that for proper function, Id3/d7 requires additional sequences to function properly. These sequences may exist within or outside of the intergenic region. The inability of the Id1/d2 element to reproduce the entire domain of endogenous visceral arch *Dlx* expression suggests the requirement for extra sequences for optimal function as well. One possibility is that these intergenic elements may require to interact with the specific promoters of the *Dlx* genes.

Interestingly the two enhancers, Id1/d2 and Id4/d6i, which are able to recapitulate sub-domains of the expression patterns of the endogenous *Dlx* genes which flank them in the genome, are also the sites of cross-regulation by the *Dlx* genes (Zerucha et al., 1999), and this study). Previously we have also described the ability of Dlx proteins to activate transcription through the 1.7 kb region from upstream the *dlx4* gene (Zerucha et al., 1997). This suggests that auto- and cross-regulatory interactions significantly contribute to *Dlx* expression during development. The importance of cross-regulatory interactions between the *Dlx* genes was probably best illustrated by the inactivation of the Id4/d6i enhancer in mice lacking the normal functions of *Dlx1* and *Dlx2* (Zerucha et al., 1999). The ability of Dlx proteins to activate transcription through these two defined elements also presents the possibility to compare the sequences to ascertain if there is any sequence similarity which may represent a recognition site for Dlx proteins. In the characterizations of each of these elements, regions were identified of approximately 110-130 bp (328-440 of Id1/d2 and 187-316 of Id4/d6i) that could be bound and activated by Dlx2. The region of Id1/d2 from 412-422 (CTATAATTTTC, Fig. 4.1B) shares >90%

similarity, or ten of eleven nucleotides, with a region from 271-281 (CTGTAATTTTC) of Id4/d6i. This sequence was the region conserved to the greatest degree between the two elements and in addition contains an ATTA/TAAT core sequence, which has been previously demonstrated to be a common motif recognized by many homeodomain proteins. The Dlx2 protein is unable to bind a double stranded oligonucleotide pair containing this element. Furthermore, the mutation of this site to CTCTAGATTTC in Id4/d6i, which removes the ATTA/TAAT core sequence, does not affect the ability of Dlx proteins to activate transcription through this element (Zerucha and Ekker unpublished observations).

The Function Of Id1/d2 And Id4/d6 Seems To Be Similar In Mice And *Drosophila*

The conservation of activity of the Id4/d6 and Id1/d2 elements between vertebrates and *Drosophila* is highly significant. Of particular interest is that these elements seem to be able to direct expression in analogous regions of the *Drosophila* and murine embryos. The Id4/d6 elements direct expression to the ventral forebrain of vertebrates and the brain of *Drosophila* during development. The forebrain has been considered a vertebrate innovation but recently the expression of a *Distal-less* gene in the developing brain of the cephalochordate amphioxus was used to suggest an earlier origin for this structure (Holland et al., 1996). If the expression we describe in this study does in fact represent a true domain of expression of the *Drosophila Dll* gene, these conclusions on the origins of the forebrain may have to be re-thought. It would be possible, then, that *Dll*-like genes have been expressed in the brains and brain-like

structures of organisms whose lineages arose prior to the vertebrate/invertebrate divergence. This could suggest that the origin of the forebrain resulted from an exaptation of pre-existing mechanisms including *Dll* expression. It is possible that vertebrate and invertebrate *Dll* brain expression arose independently of one another, but the conservation of function of the Id4/d6 enhancers seems to refute this. It can not be ruled out that the expression we observe directed by Id4/d6 is simply Id4/d6 responding solely to the endogenous *Drosophila Dll* as we have demonstrated that this element does respond to vertebrate *Dlx* genes (Zerucha et al., 1999). However the lack of reporter expression in other regions where *Dll* is expressed, most notably the limbs, implies this to be unlikely.

The Id1/d2 element directs expression to the visceral arches of vertebrates and the ventral maxillary of *Drosophila* during development. It is tempting to speculate that this might imply a common origin for vertebrate jaws and invertebrate mouth parts. However while it is possible *Dll* may have been expressed in the developing, primitive, external feeding apparatus associated with the mouth of the ancient ancestors to vertebrates and invertebrates, it seems highly unlikely that the vertebrate jaw and *Drosophila* segmented mouth appendages could have diverged from common ancestral structures. The shared pattern of expression directed by Id1/d2 and the Drosophila ETD6 elements, however, may shed light on the differences in the developmental pathways resulting in jaws or mouth appendages. The important aspect of this expression is that in *Drosophila* ETD6 is regulated by at least one of the *HOM-C* genes, while in vertebrates it appears that Id1/d2 expression is independent of the actions of the orthologous *Hox* genes. This

suggests, then, that the evolution of the development of the *Drosophila* mouth appendages has occurred to be directly dependant on patterning conferred by the *HOM-C* genes while in vertebrates this appears to not be the case. This is consistent with the observations that while *Hox* genes are expressed in the hindbrain during vertebrate development, they are not expressed in the migrating neural crest of the anterior-most visceral arch, which gives rise to the jaw structures. Comparison of Id1/d2 with the *Drosophila Dll* locus in cross-species Southern hybridizations revealed no similar regions, suggesting that any nucleotide sequence conservation between the two elements is low. In addition the ability of these elements to be regulated by the *Dlx* genes has not been conserved as Dlx proteins can not activate transcription through the ETD6 element (Zerucha and Ekker, unpublished observations). The degree of functional conservation of these elements is also unknown as transgenic mice carrying the ETD6 element do not exhibit any expression of the transgene (Hatch and Ekker, unpublished observations). It is possible that this lack of reporter expression may be due to the lack of important regions outside the ETD6 element proper, although the exact reasons why it is inactive in mice while Id1/d2 is active in *Drosophila* is not exactly clear. It is possible that *Drosophila* have evolved a different complement of requirements for optimal function of the ETD6 element such as specific interactions with other *cis*- or *trans*-acting elements. It is also possible that for similar reasons it is easier for an element to function in a less complex setting than a more complex one. This may be due to a few factors that have been conserved between a less complex and a more complex organism being sufficient to activate a functionally conserved element between two organisms of differing complexities. However in the organism of higher complexity, the hierarchies of controls

on the conserved element may have been evolutionary built up such that all the components required for its coordinated expression simply can not regulate the element from the organism of lower complexity.

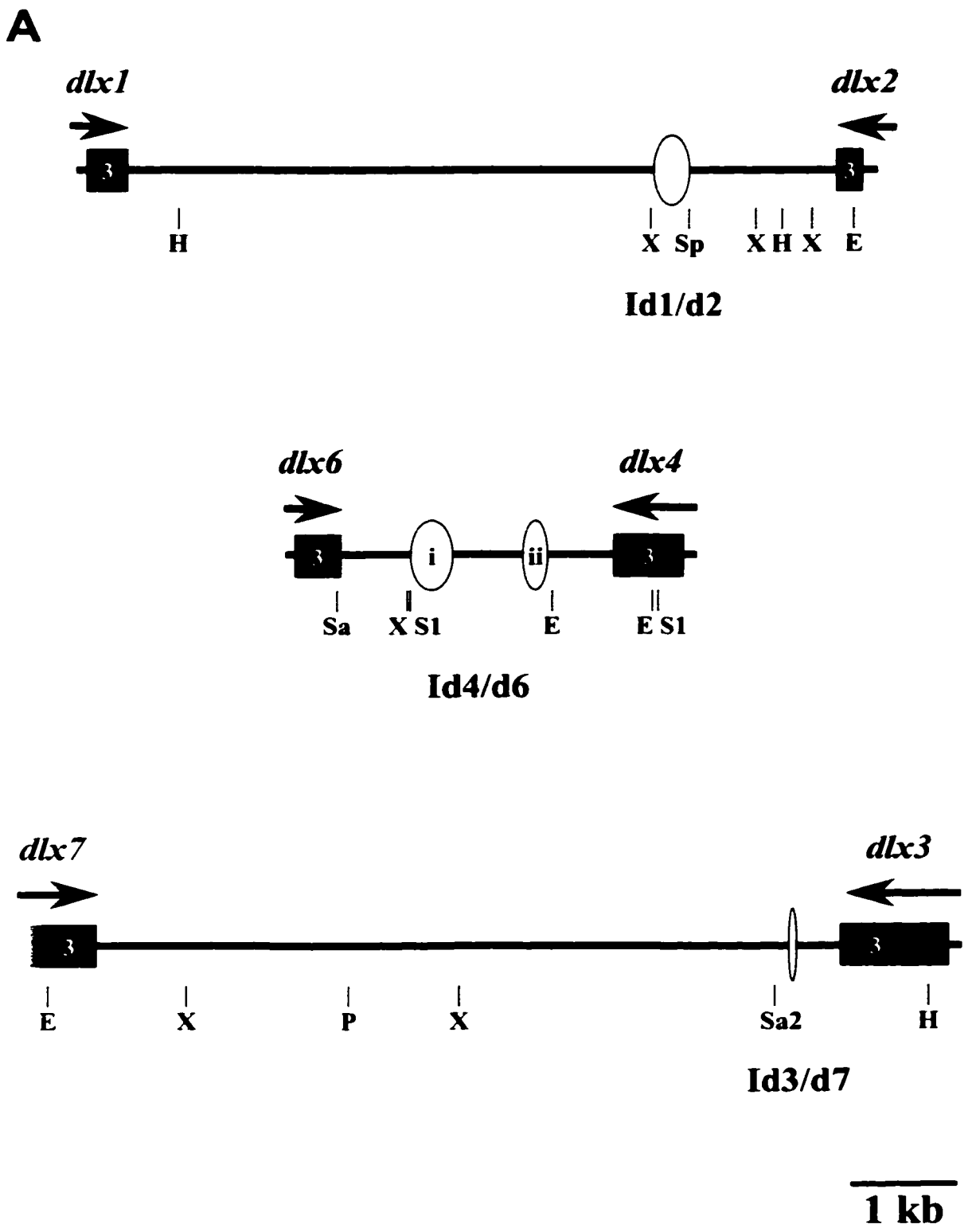
A Proposed Model For The Evolution Of The *Dlx* Gene Family

Based on the expression patterns directed by the *Dlx* intergenic elements in *Drosophila* and vertebrates we can perhaps suggest a mechanism by which this family evolved (Fig. 4.5). An ancestral *Distal-less*-like gene had a number of downstream enhancer elements associated with it, such as is proposed to exist in *Drosophila*. This gene underwent a tandem duplication event resulting in an ancestral convergently transcribed gene pair as is observed in modern urochordates such as the ascidian *Ciona intestinalis*. We propose that this ancestral pair of *Distal-less* genes, at least originally, contained a full or close to full complement of the downstream enhancer elements which are now localized to the intergenic region separating the gene pair. A series of large-scale genomic events resulted in the duplication of this ancestral pair resulting in three pairs of convergently transcribed gene pairs, which initially would have the full complement of intergenic enhancer elements. For these gene pairs to be maintained some selective pressure must be exerted on these gene pairs. We propose that one such selective pressure resulted from the loss of elements from pairs of genes. Thus while the whole gene family still had a full, or close to full complement of elements, individual gene pairs became associated with specific regions of the developing embryo such as *dlx1*/2 with the visceral arches and *dlx4*/6 with the forebrain. The function of the Id3/d7

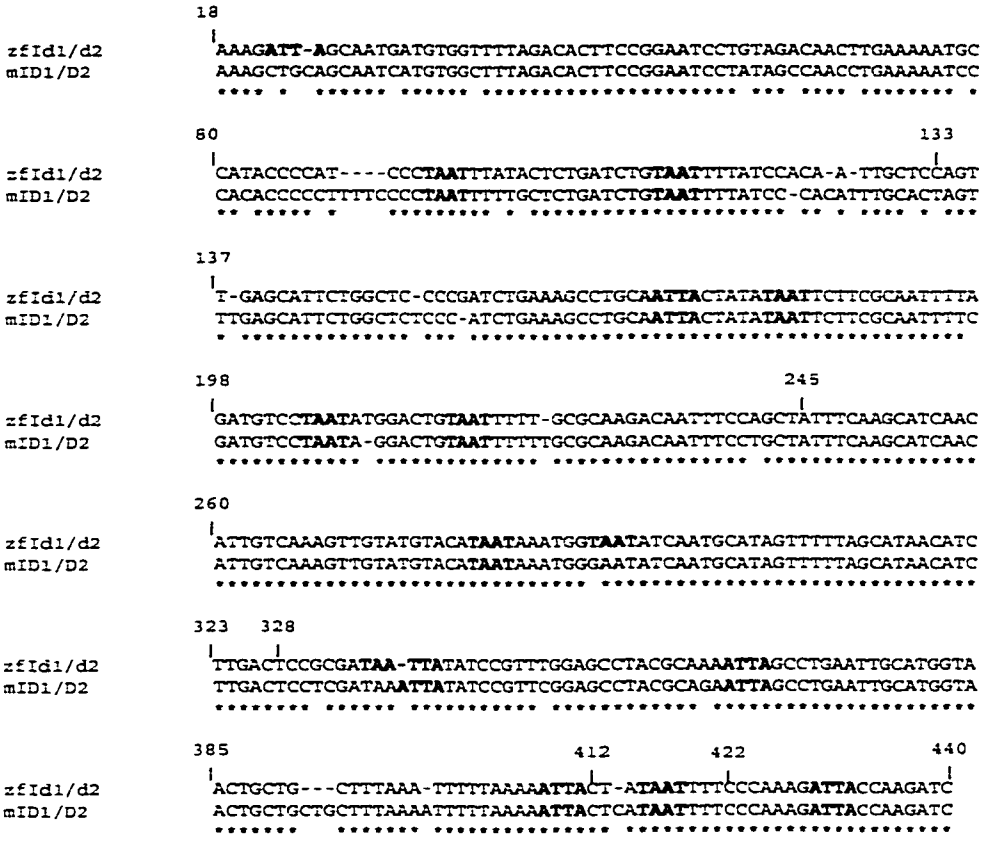
element in *Drosophila* remains to be determined, however it is known that *Drosophila Dll* is expressed in the genital disc during *Drosophila* development. It will be interesting as well to determine if the entire intergenic region from the single *C. intestinalis Dll* gene pair can similarly direct expression in transgenic *Drosophila*.

This mechanistic model does not preclude the possibility that the enhancers have taken on new functions during vertebrate evolution. However it is possible that they still have some basal activity in *Drosophila* perhaps reminiscent of their ancestral function and indicated by the activity of the vertebrate elements in *Drosophila* but not the reverse. It is interesting that if this model is true, it suggests that the biological cost of maintaining a higher number of genes than is minimally required can be seen as an investment in evolutionary potential. Essentially, the mechanism of this model provides a greater number of genes for the mechanisms of evolution to act upon to move toward greater complexity.

Figure 4.1. Identification of sequences in the intergenic regions of the paired zebrafish *dlx* genes that are highly conserved between mice and zebrafish. A) Genomic organization of the paired zebrafish dlx genes. The third exons of each of the genes is represented by a filled box and labeled appropriately. Direction of transcription of the dlx genes is indicated by an arrow. The conserved elements are shown as open ovals and are labeled. E, *Eco*RI; H, *Hind*III; P, *Pst*I; Sa, *Sal*I; Sp, *Sph*I; S1, *Sac*I; S2, *Sac*II; X, *Xho*I. B) Alignment of the Id1/d2-ID1/D2 sequences from zebrafish and mouse. Nucleotide identity between the two orthologous sequences is indicated by a star below the nucleotide. Putative homeodomain recognition sites as characterized by a TAAT/ATTA core nucleotide sequence are shaded. Numbering is indicated for the zebrafish sequence and is relative to the *Sph*I site (position 1) and *Xho*I site (position 440). The positions of the oligonucleotide primers used to construct deletion clones of the zebrafish element are indicated. C) Alignment of the Id3/d7-ID3/D7 sequences from zebrafish, mouse and human. Nucleotide sequence identity across the three species is indicated by a star below the nucleotide. The single TAAT/ATTA core sequence is shaded and numbering is relative to the zebrafish sequence.


A
dlx1
dlx2
H
X Sp
X H X E
Id1/d2
dlx6
dlx4
i
ii
Sa
X S1
E
E S1
Id4/d6
dlx7
dlx3
E
X
P
X
Sa2
H
Id3/d7
1 kb

B


18
zfId1/d2 AAAGATT-AGCAATGATGTGGTTTTAGACACTTCCGGAATCCTGTAGACAACTTGAAAAATGC
mID1/D2 AAAGCTGCAGCAATCATGTGGCTTTAGACACTTCCGGAATCCTATAGCCAACCTGAAAAATCC
80 133
zfId1/d2 CATACCCCAT----CCCTAATTTATACTCTGATCTGTAATTTTATCCACA-A-TTGCTCCAGT
mID1/D2 CACACCCCCTTTTCCCCTAATTTTTGCTCTGATCTGTAATTTTATCC-CACATTTGCACTAGT
137
zfId1/d2 T-GAGCATTCTGGCTC-CCCGATCTGAAAGCCTGCAATTACTATATAATTCTTCGCAATTTTA
mID1/D2 TTGAGCATTCTGGCTCTCCC-ATCTGAAAGCCTGCAATTACTATATAATTCTTCGCAATTTTC
198 245
zfId1/d2 GATGTCCTAATATGGACTGTAATTTTT-GCGCAAGACAATTTCCAGCTATTTCAAGCATCAAC
mID1/D2 GATGTCCTAATA-GGACTGTAATTTTTTGCGCAAGACAATTTCCTGCTATTTCAAGCATCAAC
260
zfId1/d2 ATTGTCAAAGTTGTATGTACATAATAAATGGTAATATCAATGCATAGTTTTTAGCATAACATC
mID1/D2 ATTGTCAAAGTTGTATGTACATAATAAATGGGAATATCAATGCATAGTTTTTAGCATAACATC
323 328
zfId1/d2 TTGACTCCGCGATAA-TTATATCCGTTTGGAGCCTACGCAAAATTAGCCTGAATTGCATGGTA
mID1/D2 TTGACTCCTCGATAAATTATATCCGTTCGGAGCCTACGCAGAATTAGCCTGAATTGCATGGTA
385 412 422 440
zfId1/d2 ACTGCTG---CTTTAAA-TTTTTAAAAATTACT-ATAATTTTCCCAAAGATTACCAAGATC
mID1/D2 ACTGCTGCTGCTTTAAAATTTTTAAAAATTACTCATAATTTTCCCAAAGATTACCAAGATC

C

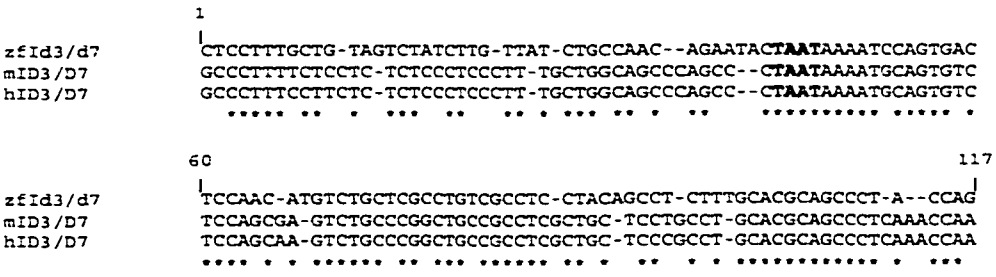

1
zfId3/d7 CTCCTTTGCTG-TAGTCTATCTTG-TTAT-CTGCCAAC--AGAATACTAATAAAATCCAGTGAC
mID3/D7 GCCCTTTTCTCCTC-TCTCCCTCCCTT-TGCTGGCAGCCCAGCC--CTAATAAAATGCAGTGTC
hID3/D7 GCCCTTTCCTTCTC-TCTCCCTCCCTT-TGCTGGCAGCCCAGCC--CTAATAAAATGCAGTGTC
60 117
zfId3/d7 TCCAAC-ATGTCTGCTCGCCTGTCGCCTC-CTACAGCCT-CTTTGCACGCAGCCCT-A--CCAG
mID3/D7 TCCAGCGA-GTCTGCCCGGCTGCCGCCTCGCTGC-TCCTGCCT-GCACGCAGCCCTCAAACCAA
hID3/D7 TCCAGCAA-GTCTGCCCGGCTGCCGCCTCGCTGC-TCCCGCCT-GCACGCAGCCCTCAAACCAA

Figure 4.2. The zebrafish Id1/d2 and Id4/d6 elements are able to direct expression in transgenic mice in domains which reflect a subset of the full expression patterns of the Dlx genes which flank their murine homologues. A) Id1/d2 directs expression of *LacZ* to the mandibular arch (arrow) in stage E11 transgenic mouse embryos. B) Id4/d6 directs expression of *LacZ* to the ventral forebrain (arrows) and the olfactory placode (o) in stage E11 transgenic mouse embryos. C) The entire *dlx3/dlx7* intergenic region directs expression of *LacZ* to the genital ridge in stage E11 transgenic mouse embryos. D) Id1/d2 directs expression to the ventral maxillary (arrow) in stage 12 transgenic *Drosophila* embryos. E) Id4/d6 directs expression to the developing brain (arrows) of stage 11 transgenic *Drosophila* embryos and F) the peripheral nervous system of stage 10 transgenic *Drosophila* embryos.

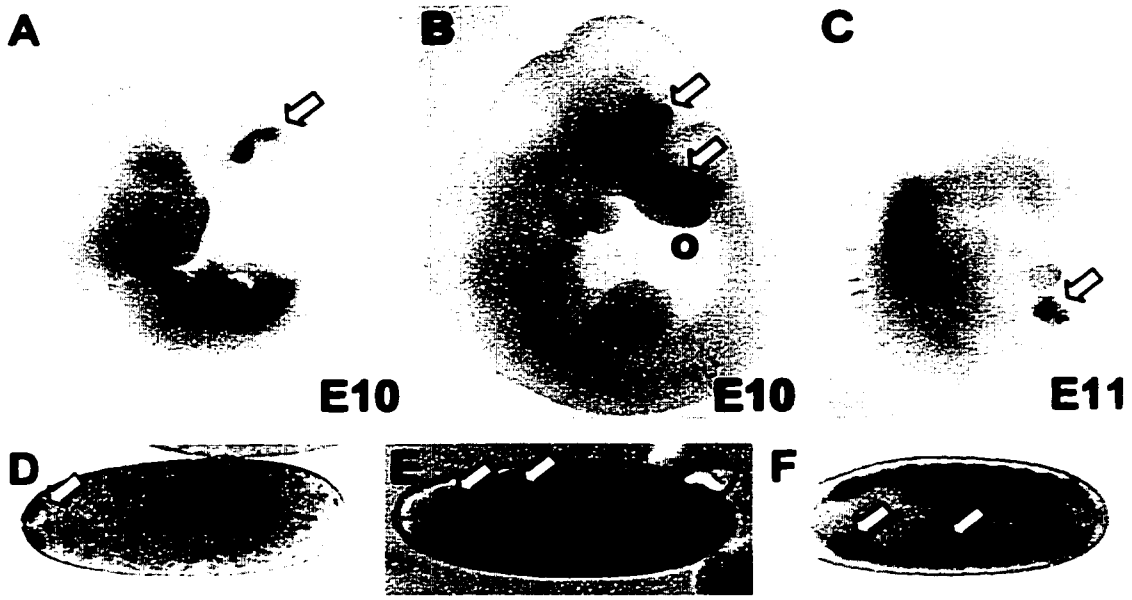

A
B
C
E10
E10
E11
D
E
F

Figure 4.3. Dlx proteins can activate transcription through a specific region of Id1/d2. A) Summary of deletion fragments of Id1/d2 in the pBLCAT2 reporter construct. The numerical names of the full length Id1/d2 element and the deletion fragments of it reflect the nucleotide sequence boundaries within the Id1/d2 element and correspond to the numbering in Fig. 4.1. The orientation of all the fragments relative to the Thymidine Kinase basal promoter (tk) and the Chloramphenicol Acetyl Transferase (CAT) gene is as shown. The direction of transcription of the CAT gene is indicated by an arrow. B) Co-transfected Dlx2 activates transcription through the full length (1-440) Id1/d2. Specifically Dlx2 is able to activate transcription through the 133-440 and 328-440 deletion fragments, but not the 1-245 and 133-323 deletion fragments. The values shown represent three independent experiments and the error bars indicate standard error on the mean.

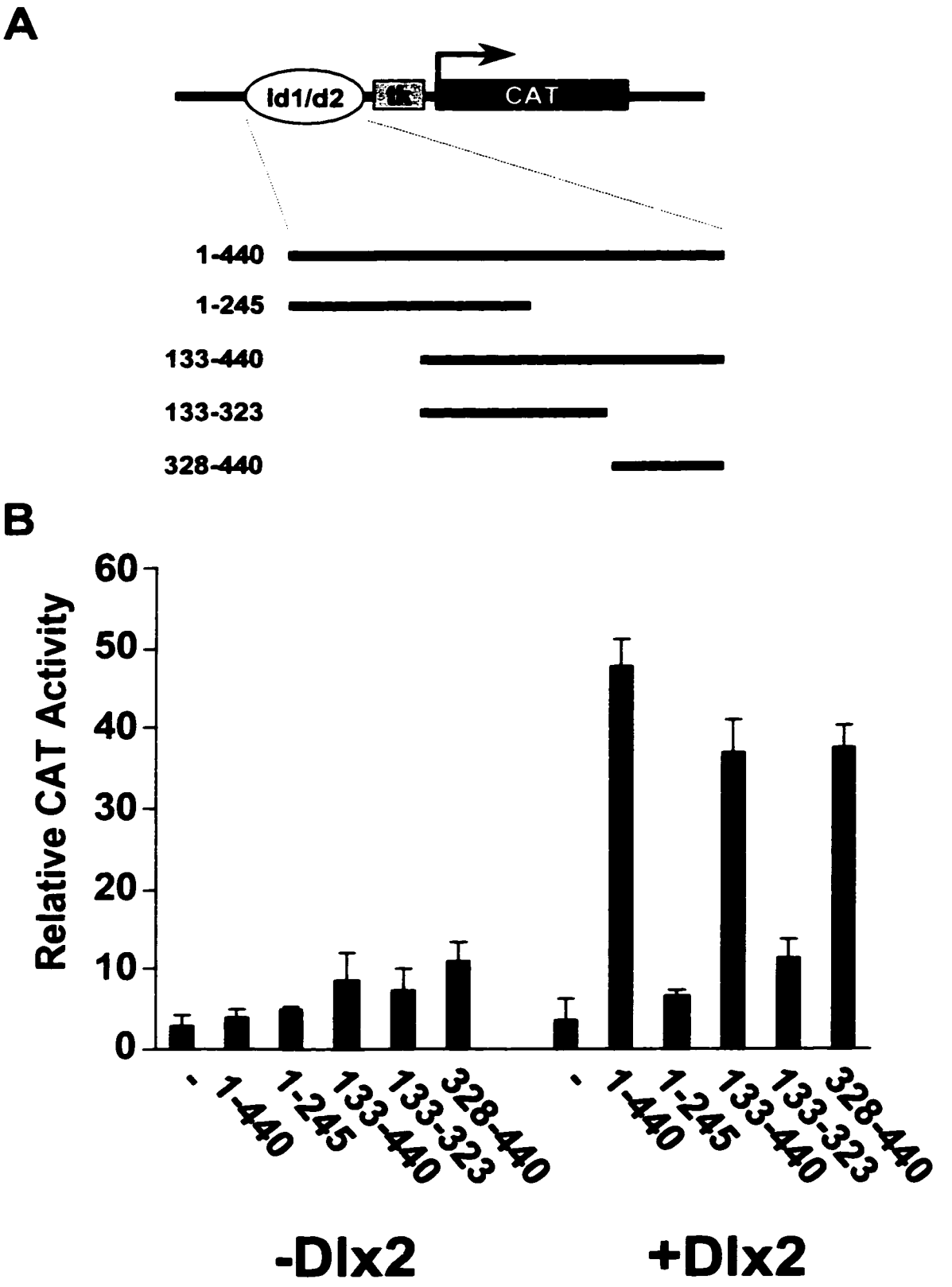

A
Id1/d2
tk
CAT
1-440
1-245
133-440
133-323
328-440
B
Relative CAT Activity
60
50
40
30
20
10
0
-
1-440
1-245
133-440
133-323
328-440
-
1-440
1-245
133-440
133-323
328-440
-Dlx2
+Dlx2

Figure 4.4. The zebrafish Dlx2 protein binds the 328-440 fragment of Id1/d2 as shown by electrophoretic mobility shift assay. The 328-440 fragment was radiolabeled and incubated with nuclear extract from the SF7-MTG-Dlx2 cell line or from SF7 cell line as a control. The filled arrow indicates the mobility shift resulting from Dlx2 interacting with the 328-440 fragment. This shift is super-shifted (open arrow) in the presence of the 9E-10 monoclonal antibody to the MTG epitope of MTG-Dlx2.

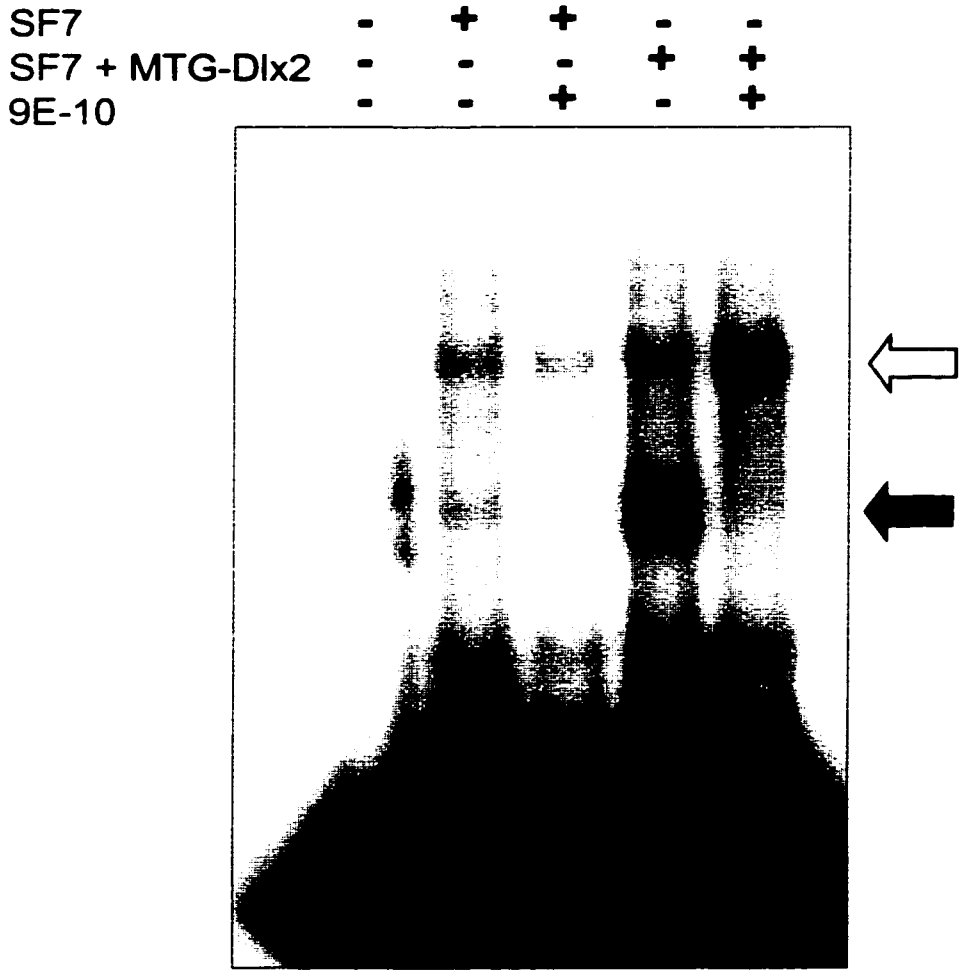
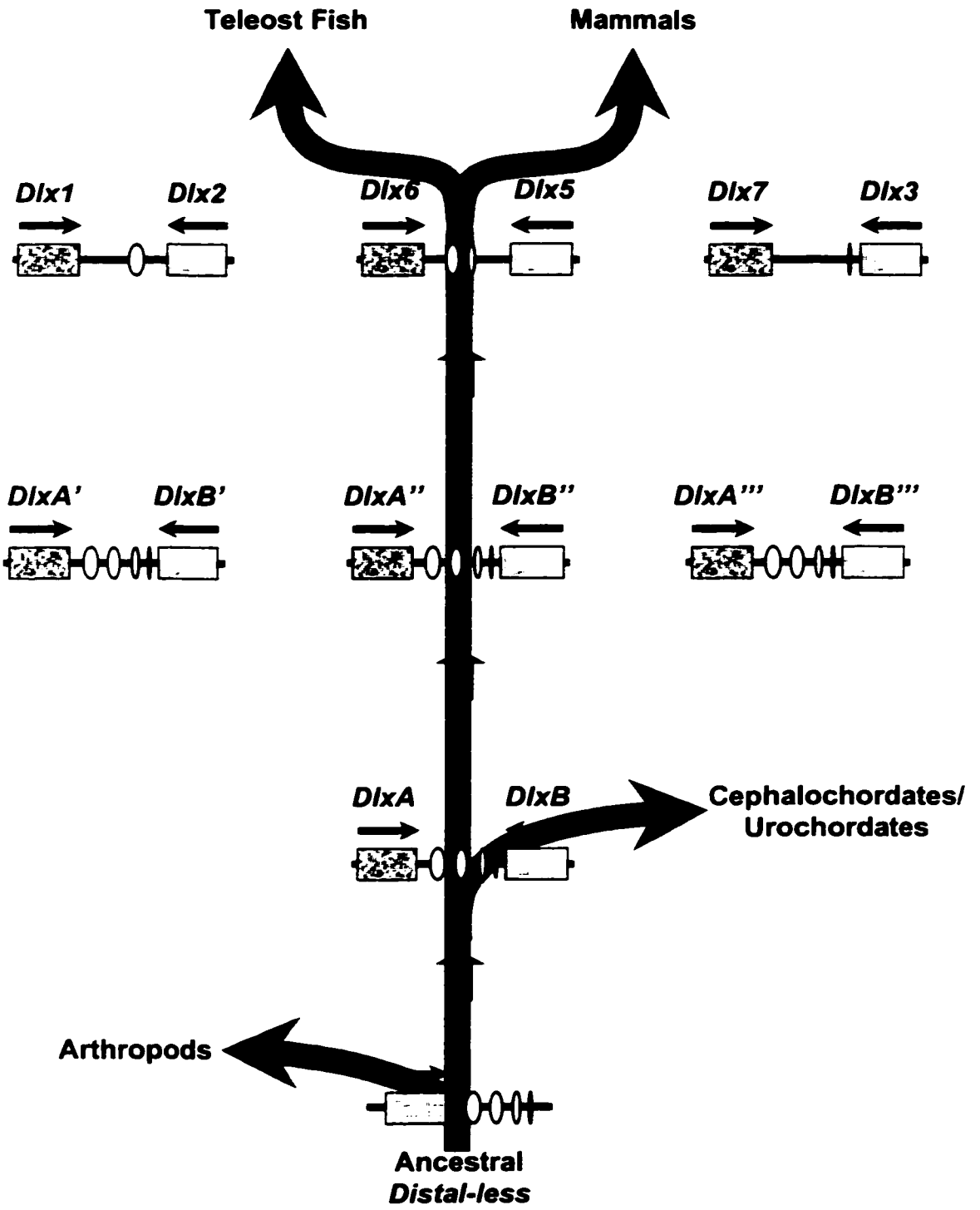

SF7 - + + - -
SF7 + MTG-Dlx2 - - - + +
9E-10 - - + - +

Figure 4.5. A proposed model for the evolution of the *Distal-less*-related gene family under the selective pressure of enhancer duplication followed by the subsequent loss of specific enhancers from different gene pairs. Genes are represented by filled boxes and their direction of transcription indicated by arrows above each gene. Enhancer elements are represented by ovals.


Teleost Fish
Mammals
Dlx1
Dlx2
Dlx6
Dlx5
Dlx7
Dlx3
DlxA'
DlxB'
DlxA''
DlxB''
DlxA'''
DlxB'''
DlxA
DlxB
Cephalochordates/
Urochordates
Arthropods
Ancestral
Distal-less

5. CONCLUSIONS

The over-riding theme of this thesis has been the importance of auto- and cross-regulatory interactions between the *Dlx* family of homeobox-containing genes and how the elements that are the sites of these interactions may have evolved. There are precedents for such auto- and cross-regulatory interactions between genes that encode transcriptional regulatory proteins, most notable the *Hox* genes that have already been discussed. A possible benefit for the evolution of auto-regulatory mechanisms, particularly for transcriptional activators is, intuitively, simply to maintain their expression once they are activated. Often during development it appears that the expression of genes encoding transcription factors is activated by gradients of extracellular signals. The optimal conditions then for activation of a gene may be fleeting, thus the ability to maintain its expression, once activated would seem to be an ability with a heavy positive selective pressure attached. This then also implies the importance of mechanisms to turn off the expression of such auto-regulatory factors. One possibility in the case of *Dlx* genes, involves inhibitory interactions with other factors such as members of the Msx family of homeodomain proteins which has been discussed previously in more detail in the Introduction.

The initial evolution of an auto-regulatory mechanism also presents an interesting possibility related to the work and findings described in this thesis. Hypothetically, I will begin with the evolution of an auto-regulatory element associated with a specific gene that encodes a transcription factor. A number of genomic events occur which result in a

number of duplications of the locus of this gene, including the auto-regulatory element. There will be little, or no selective pressure to maintain a number of identical genetic loci, therefore copies of loci will either become inactive via deleterious mutations, or will take on new functions that will somehow increase the fitness of the organism. To take on new functions, the coding region of the gene may change, the elements that direct its expression may change, or some combination of these may occur. An interesting prospect arising from this latter possibility is the evolution of a cross-regulatory element. As changes occur within the coding region of the gene, the protein that it encodes may take on new functions while retaining some of the original functions such as DNA binding. Thus a family arises which can bind the same or very similar recognition elements, but because there have been modifications outside the DNA binding domain, different members of the family may play different roles when bound to similar recognition elements. Also, changes in the regulation of the different family members will alter which specific members will even be available to interact with specific elements.

What then do the regulatory elements I describe here for the *Dlx* family represent? From the results described in chapter two, the upstream region of *dlx4* seems to represent a cross-regulatory element regulated by *dlx3*. Expression of *dlx3* precedes that of *dlx4* in the same cells of the otic vesicle during zebrafish development. Also in cultured cells *dlx3* can activate transcription through a 1.7 kb region from upstream *dlx4*. However from what we have seen in chapters three and four, it seems that most Dlx proteins are able to activate transcription from identical nucleotide sequences. Thus despite the

temporal relationship between *dlx3* and *dlx4* expression it is difficult to state with absolute certainty that this upstream region of *dlx4* represents a cross-regulatory element.

Similar conclusions can also be reached for the Id1/d2 element. I have shown in chapter four that Dlx proteins can activate transcription through this element and that at least one of them, Dlx2, can bind it. It will be interesting to see the activity of this element in mice lacking the *Dlx1* and *Dlx2* genes as this will give us a better insight as to whether it s activity is dependant on auto- or cross-regulatory interactions. It does appear that the *Dlx2* gene is the first family member to be expressed in the visceral arches during development so this may suggest that we are observing an auto-regulatory mechanism. Conversely, the loss of activity of the Id4/d6 element in mice lacking functional *Dlx1* and *Dlx2* suggests a cross-regulatory interaction.

An important question also arises from the pattern of expression directed by Id4/d6. It appears that although the endogenous murine *Dlx5* and *Dlx6* genes are roughly co-expressed in the forebrain coincident with the pattern directed by the Id4/d6 or ID5/D6 elements, the expression directed by this element more closely resembles that of *Dlx5*. In addition as previously mentioned, expression of the *Dlx2* gene precedes that of *Dlx1* and in fact all *Dlx* genes in the visceral arches (where Id1/d2 directs expression). Does this mean that these elements are acting uni-directionally? Because of the organization of the convergently transcribed, paired *Dlx* genes, the intergenic elements are located downstream of each of the genes that flank them. Thus their organization relative to the transcriptional start sites of either the flanking genes is the same. It will be

interesting to determine if insulating sequences exist within the intergenic regions or if specific interactions occur between the intergenic enhancers and specific promoters or other *cis*-regulatory elements to affect the transcription of only one gene in each *Dlx* pair.

Future work on this project may also shed light on the mechanism of evolution of cis-regulatory elements. As was mentioned in chapter four, there is no striking similarity between the vertebrate *Dlx* intergenic elements and the single known homologous region in *C. intestinalis*. It will be of great interest to determine if there is any sequence conservation between the *C. intestinalis Dll* intergenic region and that of other ascidians. Also if any of the ascidian *Dll* intergenic regions are able to direct expression in transgenic *Drosophila*, in vertebrates and of course themselves, the comparisons and contrasts with the already identified vertebrate elements will perhaps contribute to a better undertanding of the evolution of *cis*-regulatory elements as well as helping us to understand the differences between functional conservation and actual nucleotide sequence conservation.

My project has contributed to a better understanding of the transcriptional regulation of the *Dlx* family of homeobox-containing genes. My work has also contributed to an explanation for the maintenance of the close linkage of the paired *Dlx* genes and perhaps suggested a model for the evolution of this gene family. This overall project has grown considerably in the last few years and it will be of great interest to see how it evolves in the future.

6. REFERENCES CITED

Akimenko, M. A., Ekker, M., Wegner, J., Lin, W., and Westerfield, M. (1994). Combinatorial expression of three zebrafish genes related to *distal-less*: part of a homeobox gene code for the head. J Neurosci *14*, 3475-86.

Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y. L., Westerfield, M., Ekker, M., and Postlethwait, J. H. (1998). Zebrafish *hox* clusters and vertebrate genome evolution. Science *282*, 1711-4.

Anderson, S. A., Qiu, M., Bulfone, A., Eisenstat, D. D., Meneses, J., Pedersen, R., and Rubenstein, J. L. (1997a). Mutations of the homeobox genes *Dlx-1* and *Dlx-2* disrupt the striatal subventricular zone and differentiation of late born striatal neurons. Neuron *19*, 27-37.

Anderson, S. A., Eisenstat, D. D., Shi, L., and Rubenstein, J. L. (1997b). Interneuron migration from basal forebrain to neocortex: dependence on *Dlx* genes. Science *278*, 474-6.

Andrews, N. C., and Faller, D. V. (1991). A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. Nucleic Acids Res *19*, 2499.

Arcioni, L., Simeone, A., Guazzi, S., Zappavigna, V., Boncinelli, E., and Mavilio, F. (1992). The upstream region of the human homeobox gene *HOX3D* is a target for regulation by retinoic acid and HOX homeoproteins. EMBO J. *11*, 265-272.

Asano, M., Emori, Y., Saigo, K., and Shiokawa, K. (1992). Isolation and characterization of a *Xenopus* cDNA which encodes a homeodomain highly homologous to *Drosophila Distal-less*. J Biol Chem *267*, 5044-7.

Balling, R., Mutter, G., Gruss, P., and Kessel, M. (1989). Craniofacial abnormalities induced by ectopic expression of the homeobox gene *Hox-1.1* in transgenic mice. Cell *58*, 337-47.

Bateson, W. (1894). Materials for the Study of Variation Treated with Special Regard to Discontinuity in the Origin of Species (London: Macmillan and Co.).

Beachy, P. A., Krasnow, M. A., Gavis, E. R., and Hogness, D. S. (1988). An Ultrabithorax protein binds sequences near its own and the *Antennapedia* P1 promoters. Cell *55*, 1069-81.

Beauchemin, M., and Savard, P. (1992). Two *distal-less* related homeobox-containing genes expressed in regeneration blastemas of the newt. Dev Biol *154*, 55-65.

Beckers, J., Gérard, M., and Duboule, D. (1996). Transgenic analysis of a potential *Hoxd-11* limb regulatory element present in tetrapods and fish. Dev Biol *180*, 543-53.

Bodner, M., Castrillo, J. L., Theill, L. E., Deerinck, T., Ellisman, M., and Karin, M. (1988). The pituitary-specific transcription factor GHF-1 is a homeobox-containing protein. Cell *55*, 505-18.

Bridges, C. B., and Morgan, T. H. (1923). The third chromosome group of mutant characters of *Drososphila melanogaster*. Carnegie Institute of Washington Publications, 93.

Bulfone, A., Kim, H. J., Puelles, L., Porteus, M. H., Grippo, J. F., and Rubenstein, J. L. (1993a). The mouse *Dlx-2* (*Tes-1*) gene is expressed in spatially restricted domains of the forebrain, face and limbs in midgestation mouse embryos [published erratum appears in Mech Dev 1993 Aug;42(3):187]. Mech Dev *40*, 129-40.

Bulfone, A., Puelles, L., Porteus, M. H., Frohman, M. A., Martin, G. R., and Rubenstein, J. L. (1993b). Spatially restricted expression of *Dlx-1*, *Dlx-2* (*Tes-1*), *Gbx-2*, and *Wnt- 3* in the embryonic day 12.5 mouse forebrain defines potential transverse and longitudinal segmental boundaries. J Neurosci *13*, 3155-72.

Carroll, S. B., Gates, J., Keys, D. N., Paddock, S. W., Panganiban, G. E., Selegue, J. E., and Williams, J. A. (1994). Pattern formation and eyespot determination in butterfly wings. Science *265*, 109-14.

Clerc, R. G., Corcoran, L. M., LeBowitz, J. H., Baltimore, D., and Sharp, P. A. (1988). The B-cell-specific Oct-2 protein contains POU box- and homeobox-type domains. Genes Dev *2*, 1570-81.

Cohen, S. M., Bronner, G., Kuttner, F., Jurgens, G., and Jackle, H. (1989). *Distal-less* encodes a homoeodomain protein required for limb development in *Drosophila*. Nature *338*, 432-4.

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

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

Cohen, B., Simcox, A. A., and Cohen, S. M. (1993). Allocation of the thoracic imaginal primordia in the *Drosophila* embryo. Development *117*, 597-608.

Cooke, J., Nowak, M. A., Boerlijst, M., and Maynard-Smith, J. (1997). Evolutionary origins and maintenance of redundant gene expression during metazoan development. Trends Genet *13*, 360-4.

Cormack, B. P., Valdivia, R. H., and Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). Gene *173*, 33-8.

Desplan, C., Theis, J., and O'Farrell, P. H. (1985). The *Drosophila* developmental gene, *engrailed*, encodes a sequence-specific DNA binding activity. Nature *318*, 630-5.

Desplan, C., Theis, J., and O'Farrell, P. H. (1988). The sequence specificity of homeodomain-DNA interaction. Cell *54*, 1081-90.

deVries, C. J. M., Boer, J. d., Joore, J., Strahle, U., Achterberg, T. A. E. v., Hylebroeck, D., Verschueren, K., Miyazono, K., Raajj, A. J. M. v. d. E.-v., and Zivkovic, D. (1996). Active Complex formation of type I and type II activin and TGF beta receptors in vivo as studied by overexpression in zebrafish embryos. Mech Dev *54*, 225-236.

Di Gregorio, A., Spagnuolo, A., Ristoratore, F., Pischetola, M., F, Aniello, Branno, M., Cariello, L., and DiLauro, R. (1995). Cloning of ascidian homeobox genes provides evidence for a primordial chordate cluster. Gene *156*, 253-257.

Dirksen, M. L., Mathers, P., and Jamrich, M. (1993). Expression of a *Xenopus Distal-less* homeobox gene involved in forebrain and cranio-facial development. Mech Dev *41*, 121-8.

Dollé, P., Price, M., and Duboule, D. (1992). Expression of the murine *Dlx-1* homeobox gene during facial, ocular and limb development. Differentiation *49*, 93-9.

Duboule, D. (1994). Guidebook to the Homeobox Genes (Oxford: Oxford University Press).

Duboule, D. (1998). *Hox* is in the hair: a break in colinearity? Genes Dev *12*, 1-4.

Ekker, M., Akimenko, M. A., Bremiller, R., and Westerfield, M. (1992). Regional expression of three homeobox transcripts in the inner ear of zebrafish embryos. Neuron *9*, 27-35.

Ekker, S. C., Young, K. E., von Kessler, D. P., and Beachy, P. A. (1991). Optimal DNA sequence recognition by the Ultrabithorax homeodomain of *Drosophila*. EMBO J *10*, 1179-86.

Ekker, S. C., Jackson, D. G., von Kessler, D. P., Sun, B. I., Young, K. E., and Beachy, P. A. (1994). The degree of variation in DNA sequence recognition among four *Drosophila* homeotic proteins. EMBO J *13*, 3551-60.

Ellies, D. L., Langille, R. M., Martin, C. C., Akimenko, M. A., and Ekker, M. (1997a). Specific craniofacial cartilage dysmorphogenesis coincides with a loss of *dlx* gene expression in retinoic acid-treated zebrafish embryos. Mech Dev *61*, 23-36.

Ellies, D. L., Stock, D. W., Hatch, G., Giroux, G., Weiss, K. M., and Ekker, M. (1997b). Relationship between the genomic organization and the overlapping embryonic expression patterns of the zebrafish *dlx* genes. Genomics *45*, 580-90.

Faiella, A., Zappavigna, V., Mavillo, F., and Boncinelli, E. (1994). Inhibition of retinoic acid-induced activation of 3' human *HOXB* genes by antisense oligonucleotides affects sequential activation of genes located upstream in the four *HOX* clusters. Proc. Natl. Acad. Sci. U.S.A. *91*, 5335-5339.

Fainsod, A., Bogarad, L. D., Ruusala, T., Lubin, M., Crothers, D. M., and Ruddle, F. H. (1986). The homeodomain of a murine protein binds 5' to its own homeobox. Proc. Natl. Acad. Sci. U.S.A. *83*, 9532-6.

Fang, H., and Elinson, R. P. (1996). Patterns of *distal-less* gene expression and inductive interactions in the head of the direct developing frog *Eleutherodactylus coqui*. Dev Biol *179*, 160-72.

Ferrari, D., Sumoy, L., Gannon, J., Sun, H., Brown, A. M., Upholt, W. B., and Kosher, R. A. (1995). The expression pattern of the *Distal-less* homeobox-containing gene *Dlx-5* in the developing chick limb bud suggests its involvement in apical ectodermal ridge activity, pattern formation, and cartilage differentiation. Mech Dev *52*, 257-64.

Finney, M., Ruvkun, G., and Horvitz, H. R. (1988). The *C. elegans* cell lineage and differentiation gene *unc-86* encodes a protein with a homeodomain and extended similarity to transcription factors. Cell *55*, 757-69.

Garcia-Fernandez, J., and Holland, P. W. (1994). Archetypal organization of the amphioxus *Hox* gene cluster. Nature *370*, 563-6.

Gehring, W. (1967). [Formation of a complete mid leg with sternopleura in the antennae region by the mutant Nasobemia (Ns) of *Drosophila melanogaster*]. Arch Julius Klaus Stift Vererbungsforsch Sozialanthropol Rassenhyg *41:2-4*, 44-54.

Gérard, M., Zákány, J., and Duboule, D. (1997). Interspecies exchange of a *Hoxd* enhancer in vivo induces premature transcription and anterior shift of the sacrum. Dev Biol *190*, 32-40.

Gorfinkiel, N., Morata, G., and Guerrero, I. (1997). The homeobox gene *Distal-less* induces ventral appendage development in *Drosophila*. Genes Dev *11*, 2259-71.

Gould, A., Morrison, A., Sproat, G., White, R. A., and Krumlauf, R. (1997). Positive cross-regulation and enhancer sharing: two mechanisms for specifying overlapping *Hox* expression patterns. Genes Dev *11*, 900-13.

Gould, S. J. (1977). Ontogeny and Phylogeny (Cambridge: The Belknap Press of Harvard University Press).

Gyurkovics, H., Gausz, J., Kummer, J., and Karch, F. (1990). A new homeotic mutation in the *Drosophila bithorax* complex removes a boundary separating two domains of regulation. EMBO J *9*, 2579-85.

Haerry, T. E., and Gehring, W. J. (1996). Intron of the mouse *Hoxa-7* gene contains conserved homeodomain binding sites that can function as an enhancer element in *Drosophila*. Proc Natl Acad Sci U S A *93*, 13884-9.

Haerry, T. E., and Gehring, W. J. (1997). A conserved cluster of homeodomain binding sites in the mouse *Hoxa-4* intron functions in *Drosophila* embryos as an enhancer that is directly regulated by Ultrabithorax. Dev Biol *186*, 1-15.

Hagstrom, K., Muller, M., and Schedl, P. (1996). Fab-7 functions as a chromatin domain boundary to ensure proper segment specification by the *Drosophila bithorax* complex. Genes Dev *10*, 3202-15.

Han, K., Levine, M. S., and Manley, J. L. (1989). Synergistic Activation and Repression of Transcription by *Drosophila* Homeobox Proteins. Cell *56*, 573-583.

Hérault, Y., Fraudeau, N., Zákány, J., and Duboule, D. (1997). *Ulnaless* (*Ul*), a regulatory mutation inducing both loss-of-function and gain-of-function of posterior *Hoxd* genes. Development *124*, 3493-500.

Herr, W., Sturm, R. A., Clerc, R. G., Corcoran, L. M., Baltimore, D., Sharp, P. A., Ingraham, H. A., Rosenfeld, M. G., Finney, M., and Ruvkun, G. (1988). The POU domain: a large conserved region in the mammalian *pit-1*, *oct-1*, *oct-2*, and *Caenorhabditis elegans unc-86* gene products. Genes Dev *2*, 1513-6.

Hoey, T., and Levine, M. (1988). Divergent homeobox proteins recognize similar DNA sequences in *Drosophila*. Nature *332*, 858-61.

Hogan, B., Constantini, F., and Lacy, E. (1986). Manipulating the Mouse Embryo (Cold Spring Harbor: Cold Spring Harbor Laboratory Press).

Holland, N. D., Panganiban, G., Henyey, E. L., and Holland, L. Z. (1996). Sequence and developmental expression of *AmphiDll*, an amphioxus *Distal- less* gene transcribed in the ectoderm, epidermis and nervous system: insights into evolution of craniate forebrain and neural crest. Development *122*, 2911-20.

Iler, N., Rowitch, D. H., Echelard, Y., McMahon, A. P., and Abate-Shen, C. (1995). A single homeodomain binding site restricts spatial expression of *Wnt-1* in the developing brain. Mech Dev *53*, 87-96.

Ingraham, H. A., Chen, R. P., Mangalam, H. J., Elsholtz, H. P., Flynn, S. E., Lin, C. R., Simmons, D. M., Swanson, L., and Rosenfeld, M. G. (1988). A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. Cell *55*, 519-29.

Jaynes, J. B., and O'Farrell, P. H. (1988). Activation and repression of transcription by homoeodomain-containing proteins that bind a common site. Nature *336*, 744-9.

Kapiloff, M. S., Farkash, Y., Wegner, M., and Rosenfeld, M. G. (1991). Variable effects of phosphorylation of Pit-1 dictated by the DNA response elements. Science *253*, 786-9.

Kappen, C., and Ruddle, F. H. (1993). Evolution of a regulatory gene family: *HOM/HOX* genes. Curr Opin Genet Dev *3*, 931-8.

Karch, F., Galloni, M., Sipos, L., Gausz, J., Gyurkovics, H., and Schedl, P. (1994). *Mcp* and *Fab-7*: molecular analysis of putative boundaries of *cis*- regulatory domains in the *bithorax* complex of *Drosophila melanogaster*. Nucleic Acids Res *22*, 3138-46.

Keegan, L. P., Haerry, T. E., Crotty, D. A., Packer, A. I., Wolgemuth, D. J., and Gehring, W. J. (1997). A sequence conserved in vertebrate *Hox* gene introns functions as an enhancer regulated by posterior homeotic genes in *Drosophila* imaginal discs. Mech Dev *63*, 145-57.

Kelly, G. M., Erezyilmas, D. F., and Moon, R. T. (1995a). Induction of a secondary embryonic axis in zebrafish occurs following the overexpression of *ß catenin*. Mech Dev *53*, 261-273.

Kelly, G. M., Greenstein, P., Erezyilmas, D. F., and Moon, R. T. (1995b). Zebrafish *wnt8* and *wnt8b* share a common activity but are involved in distinct developmental pathways. Development *121*, 1787-1799.

Kenyon, C. (1994). If Birds Can Fly, Why Can't We? Homeotic Genes and Evolution. Cell *78*, 175-180.

Kessel, M., Balling, R., and Gruss, P. (1990). Variations of cervical vertebrae after expression of a *Hox-1.1* transgene in mice. Cell *61*, 301-8.

Kimura, C., Takeda, N., Suzuki, M., Oshimura, M., Aizawa, S., and Matsuo, I. (1997). *Cis*-acting elements conserved between mouse and pufferfish *Otx2* genes govern the expression in mesencephalic neural crest cells. Development *124*, 3929-41.

Kissinger, C. R., Liu, B. S., Martin-Blanco, E., Kornberg, T. B., and Pabo, C. O. (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions. Cell *63*, 579-90.

Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., and Pabo, C. O. (1994). Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. Cell *77*, 21-32.

Krasnow, M. A., Saffman, E. E., Kornfeld, K., and Hogness, D. S. (1989). Transcriptional activation and repression by Ultrabithorax proteins in cultured *Drosophila* cells. Cell *57*, 1031-43.

Laughon, A., and Scott, M. P. (1984). Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. Nature *310*, 25-31.

Laughon, A., Howell, W., and Scott, M. P. (1988). The interaction of proteins encoded by *Drosophila* homeotic and segmentation genes with specific DNA sequences. Development *104*, 75-83.

Le Mouellic, H., Lallemand, Y., and Brulet, P. (1992). Homeosis in the mouse induced by a null mutation in the *Hox-3.1* gene. Cell *69*, 251-64.

Lewis, E. B. (1978). A gene complex controlling segmentation in *Drosophila*. Nature *276*, 565-70.

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

Long, Q., Meng, A., Wang, H., Jessen, J. R., Farrell, M. J., and Lin, S. (1997). *GATA-1* expression pattern can be recapitulated in living transgenic zebrafish using a *GFP* reporter gene. Development *124*, 4105-4111.

Luckow, B., and Schütz, G. (1987). CAT constructions with multiple unique restriction sites for the functional analysis of eukaryotic promoters and regulatory elements. Nucleic Acids Res *15*, 5490.

Lutz, B., Lu, H. C., Eichele, G., Miller, D., and Kaufman, T. C. (1996). Rescue of *Drosophila labial* null mutant by the chicken ortholog *Hoxb-1* demonstrates that the function of *Hox* genes is phylogenetically conserved. Genes Dev *10*, 176-84.

Malicki, J., Schughart, K., and McGinnis, W. (1990). Mouse *Hox-2.2* specifies thoracic segmental identity in *Drosophila* embryos and larvae. Cell *63*, 961-7.

Malicki, J., Bogarad, L. D., Martin, M. M., Ruddle, F. H., and McGinnis, W. (1993). Functional analysis of the mouse homeobox gene *HoxB9* in *Drosophila* development. Mech Dev *42*, 139-50.

Malicki, J., Schier, A. F., Solnica-Krezel, L., Stemple, D. L., Neuhauss, S. C., Stainier, D. Y., Abdelilah, S., Rangini, Z., Zwartkruis, F., and Driever, W. (1996). Mutations affecting development of the zebrafish ear. Development *123*, 275-83.

Marshall, H., Studer, M., Pöpperl, H., Aparicio, S., Kuroiwa, A., Brenner, S., and Krumlauf, R. (1994). A conserved retinoic acid response element required for early expression of the homeobox gene *Hoxb-1*. Nature *370*, 567-71.

McGinnis, N., Kuziora, M. A., and McGinnis, W. (1990). Human *Hox-4.2* and *Drosophila* deformed encode similar regulatory specificities in *Drosophila* embryos and larvae. Cell *63*, 969-76.

McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A., and Gehring, W. J. (1984). A conserved DNA sequence in homoeotic genes of the *Drosophila Antennapedia* and *bithorax* complexes. Nature *308*, 428-33.

McGuinness, T., Porteus, M. H., Smiga, S., Bulfone, A., Kingsley, C., Qiu, M., Liu, J. K., Long, J. E., Xu, D., and Rubenstein, J. L. (1996). Sequence, organization, and transcription of the *Dlx-1* and *Dlx-2* locus. Genomics *35*, 473-85.

Morasso, M. I., Jamrich, M., and Sargent, T. D. (1994). The homeodomain gene *Xenopus Distal-less-like-2* (*Xdll-2*) is regulated by a conserved mechanism in amphibian and mammalian epidermis. Dev Biol *162*, 267-76.

Morasso, M. I., Mahon, K. A., and Sargent, T. D. (1995). A *Xenopus distal-less* gene in transgenic mice: conserved regulation in distal limb epidermis and other sites of epithelial-mesenchymal interaction. Proc Natl Acad Sci U S A *92*, 3968-72.

Morasso, M. I., Grinberg, A., Robinson, G., Sargent, T. D., and Mahon, K. A. (1999). Placental failure in mice lacking the homeobox gene *Dlx3*. Proc Natl Acad Sci U S A *96*, 162-7.

Morrison, A., Chaudhuri, C., Ariza-McNaughton, L., Muchamore, I., Kuroiwa, A., and Krumlauf, R. (1995). Comparative analysis of chicken *Hoxb-4* regulation in transgenic mice. Mech Dev *53*, 47-59.

Mullen, L. M., Bryant, S. V., Torok, M. A., Blumberg, B., and Gardiner, D. M. (1996). Nerve dependency of regeneration: the role of Distal-less and FGF signaling in amphibian limb regeneration. Development *122*, 3487-97.

Muller, M., Affolter, M., Leupin, W., Otting, G., Wuthrich, K., and Gehring, W. J. (1988). Isolation and sequence-specific DNA binding of the Antennapedia homeodomain. EMBO J *7*, 4299-304.

Muller, M. M., Ruppert, S., Schaffner, W., and Matthias, P. (1988). A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. Nature *336*, 544-51.

Nakamura, S., Stock, D. W., Wydner, K. L., Bollekens, J. A., Takeshita, K., Nagai, B. M., Chiba, S., Kitamura, T., Freeland, T. M., Zhao, Z., Minowada, J., Lawrence, J. B., Weiss, K. M., and Ruddle, F. H. (1996). Genomic analysis of a new mammalian *distal-less* gene: *Dlx7*. Genomics *38*, 314-24.

Nonchev, S., Maconochie, M., Gould, A., Morrison, A., and Krumlauf, R. (1997). Cross-regulatory interactions between *Hox* genes and the control of segmental expression in the vertebrate central nervous system. Cold Spring Harbor Symposia On Quantitative Biology *62*, 313-23.

Nowak, M. A., Boerlijst, M. C., Cooke, J., and Smith, J. M. (1997). Evolution of genetic redundancy. Nature *388*, 167-71.

O'Hara, E., Cohen, B., Cohen, S. M., and McGinnis, W. (1993). *Distal-less* is a downstream gene of *Deformed* required for ventral maxillary identity. Development *117*, 847-56.

Ohno, S. (1970). Evolution by Gene Duplication (Heidelberg: Springer).

Otting, G., Qian, Y. Q., Billeter, M., Muller, M., Affolter, M., Gehring, W. J., and Wuthrich, K. (1990). Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. EMBO J *9*, 3085-92.

Ozcelik, T., Porteus, M. H., Rubenstein, J. L., and Francke, U. (1992). *DLX2* (*TES1*), a homeobox gene of the *Distal-less* family, assigned to conserved regions on human and mouse chromosomes 2. Genomics *13*, 1157-61.

Palopoli, M. F., and Patel, N. H. (1998). Evolution of the interaction between *Hox* genes and a downstream target. Curr Biol *8*, 587-90.

Panganiban, G., Sebring, A., Nagy, L., and Carroll, S. (1995). The development of crustacean limbs and the evolution of arthropods. Science *270*, 1363-6.

Papalopulu, N., and Kintner, C. (1993). *Xenopus Distal-less* related homeobox genes are expressed in the developing forebrain and are induced by planar signals. Development *117*, 961-75.

Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S., and Aggarwal, A. K. (1999). Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. Nature *397*, 714-719.

Peichel, C. L., Prabhakaran, B., and Vogt, T. F. (1997). The mouse *Ulnaless* mutation deregulates posterior *HoxD* gene expression and alters appendicular patterning. Development *124*, 3481-92.

Phelan, M. L., Sadoul, R., and Featherstone, M. S. (1994). Functional differences between HOX proteins conferred by two residues in the homeodomain N-terminal arm. Mol Cell Biol *14*, 5066-75.

Phelan, M. L., and Featherstone, M. S. (1997). Distinct HOX N-terminal arm residues are responsible for specificity of DNA recognition by HOX monomers and HOX-PBX heterodimers. J Biol Chem *272*, 8635-43.

Piper, D. E., Batchelor, A. H., Chang, C.-P., Cleary, M. L., and Wolberger, C. (1999). Structure of a HoxB1-Pbx1 Heterodimer Bound to DNA: Role of the Hexapeptide and a Fourth Homeodomain Helix in Complex Formation. Cell *96*, 587-597.

Pöpperl, H., and Featherstone, M. S. (1992). An autoregulatory element of the murine *Hox-4.2* gene. EMBO J *11*, 3673-80.

Pöpperl, H., Bienz, M., Studer, M., Chan, S. K., Aparicio, S., Brenner, S., Mann, R. S., and Krumlauf, R. (1995). Segmental expression of *Hoxb-1* is controlled by a highly conserved autoregulatory loop dependent upon exd/pbx. Cell *81*, 1031-42.

Porteus, M. H., Bulfone, A., Ciaranello, R. D., and Rubenstein, J. L. (1991). Isolation and characterization of a novel cDNA clone encoding a homeodomain that is developmentally regulated in the ventral forebrain [published erratum appears in Neuron 1992 Jul;9(1):187]. Neuron *7*, 221-9.

Prefontaine, G. G., Lemieux, M. E., Giffin, W., Schild-Poulter, C., Pope, L., LaCasse, E., Walker, P., and Hache, R. J. (1998). Recruitment of octamer transcription factors to DNA by glucocorticoid receptor. Mol Cell Biol *18*, 3416-30.

Price, M., Lemaistre, M., Pischetola, M., Di Lauro, R., and Duboule, D. (1991). A mouse gene related to *Distal-less* shows a restricted expression in the developing forebrain. Nature *351*, 748-51.

Qian, Y. Q., Billeter, M., Otting, G., Muller, M., Gehring, W. J., and Wuthrich, K. (1989). The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors [published erratum appears in Cell 1990 May 4;61(3):548]. Cell *59*, 573-80.

Qiu, M., Bulfone, A., Martinez, S., Meneses, J. J., Shimamura, K., Pedersen, R. A., and Rubenstein, J. L. (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*, 2523-38.

Qiu, M., Bulfone, A., Ghattas, I., Meneses, J. J., Christensen, L., Sharpe, P. T., Presley, R., Pedersen, R. A., and Rubenstein, J. L. (1997). Role of the *Dlx* homeobox genes in proximal-distal 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*, 165-84.

Robinson, G. W., Wray, S., and Mahon, K. A. (1991). Spatially restricted expression of a member of a new family of murine *Distal-less* homeobox genes in the developing forebrain. New Biol *3*, 1183-94.

Robinson, G. W., and Mahon, K. A. (1994). Differential and overlapping expression domains of *Dlx-2* and *Dlx-3* suggest distinct roles for *Distal-less* homeobox genes in craniofacial development. Mech Dev *48*, 199-215.

Rossi, E., Faiella, A., Zeviani, M., Labeit, S., Floridia, G., Brunelli, S., Cammarata, M., Boncinelli, E., and Zuffardi, O. (1994). Order of six loci at 2q24-q31 and orientation of the *HOXD* locus. Genomics *24*, 34-40.

Ryoo, H. M., Hoffmann, H. M., Beumer, T., Frenkel, B., Towler, D. A., Stein, G. S., Stein, J. L., van Wijnen, A. J., and Lian, J. B. (1997). Stage-specific expression of *Dlx-5* during osteoblast differentiation: involvement in regulation of *osteocalcin* gene expression. Mol Endocrinol *11*, 1681-94.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning A Laboratory Manual (Cold Spring Harbor: Cold Spring Harbor Laboratory Press).

Sato, T. (1984). A new homeotic mutation affecting antennae and legs. *Drosophila* Information Service *60*, 180-182.

Scheidereit, C., Cromlish, J. A., Gerster, T., Kawakami, K., Balmaceda, C. G., Currie, R. A., and Roeder, R. G. (1988). A human lymphoid-specific transcription factor that activates immunoglobulin genes is a homoeobox protein. Nature *336*, 551-7.

Scherer, S. W., Poorkaj, P., Massa, H., Soder, S., Allen, T., Nunes, M., Geshuri, D., Wong, E., Belloni, E. and Little, S. (1994). Physical mapping of the split hand/split foot locus on chromosome 7 and implication in syndromic ectrodactyly. Hum Mol Genet *3*, 1345-54.

Scherer, S. W., Heng, H. H., Robinson, G. W., Mahon, K. A., Evans, J. P., and Tsui, L. C. (1995). Assignment of the human homolog of mouse *Dlx3* to chromosome 17q21.3-q22 by analysis of somatic cell hybrids and fluorescence in situ hybridization. Mamm Genome *6*, 310-1.

Scott, M. P., and Weiner, A. J. (1984). Structural relationships among genes that control development: sequence homology between the *Antennapedia*, *Ultrabithorax* and *fushi tarazu* loci of *Drosophila*. Proc. Natl. Acad. Sci. U.S.A. *81*, 4115-4119.

Segil, N., Roberts, S. B., and Heintz, N. (1991a). Cell-cycle-regulated phosphorylation of the transcription factor Oct-1. Cold Spring Harb Symp Quant Biol *56*, 285-92.

Segil, N., Roberts, S. B., and Heintz, N. (1991b). Mitotic phosphorylation of the Oct-1 homeodomain and regulation of Oct- 1 DNA binding activity. Science *254*, 1814-6.

Seo, H. C., Drivenes, Ellingsen, S., and Fjose, A. (1998a). Expression of two zebrafish homologues of the murine *Six3* gene demarcates the initial eye primordia. Mech Dev *73*, 45-57.

Seo, H. C., Drivenes, O., Ellingsen, S., and Fjose, A. (1998b). Transient expression of a novel *Six3*-related zebrafish gene during gastrulation and eye formation. Gene *216*, 39-46.

Sharpe, J., Nonchev, S., Gould, A., Whiting, J., and Krumlauf, R. (1998). Selectivity, sharing and competitive interactions in the regulation of *Hoxb* genes. EMBO J *17*, 1788-98.

Shepherd, J. C., McGinnis, W., Carrasco, A. E., De Robertis, E. M., and Gehring, W. J. (1984). Fly and frog homeodomains show homologies with yeast mating type regulatory proteins. Nature *310*, 70-1.

Shimamura, K., Martinez, S., Puelles, L., and Rubenstein, J. L. (1997). Patterns of gene expression in the neural plate and neural tube subdivide the embryonic forebrain into transverse and longitudinal domains. Dev Neurosci *19*, 88-96.

Shirasawa, T., Sakamoto, K., and Tkahashi, H. (1994). Molecular cloning and evolutional analysis of a mammalian homologue of the *Distal-less 3* (*Dlx-3*) homeobox gene. FEBS Lett *351*, 380-4.

Simeone, A., Acampora, D., Pannese, M., D'Esposito, M., Stornaiuolo, A., Gulisano, M., Mallamaci, A., Kastury, K., Druck, T. and Huebner, K. (1994). Cloning and characterization of two members of the vertebrate *Dlx* gene family. Proc Natl Acad Sci U S A *91*, 2250-4.

Spradling, A. C., and Rubin, G. M. (1982). Transposition of cloned P elements into *Drosophila* germ line chromosomes. Science *218*, 341-7.

Stock, D. W., Ellies, D. L., Zhao, Z., Ekker, M., Ruddle, F. H., and Weiss, K. M. (1996). The evolution of the vertebrate *Dlx* gene family. Proc Natl Acad Sci U S A *93*, 10858-63.

Studer, M., Gavalas, A., Marshall, H., Ariza-McNaughton, L., Rijli, F. M., Chambon, P., and Krumlauf, R. (1998). Genetic interactions between *Hoxa1* and *Hoxb1* reveal new roles in regulation of early hindbrain patterning. Development *125*, 1025-36.

Sturm, R. A., Das, G., and Herr, W. (1988a). The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeobox subdomain. Genes Dev *2*, 1582-99.

Sturm, R. A., and Herr, W. (1988b). The POU domain is a bipartite DNA-binding structure. Nature *336*, 601-4.

Sunkel, C. E., and Whittle, J. R. S. (1987). *Brista*: a gene involved in the specification and differentiation of distal cephalic and thoracic structures in *Drosophila melanogaster*. Roux's Arch Dev Biol *196*, 124-132.

Thomas, B. L., Tucker, A. S., Qui, M., Ferguson, C. A., Hardcastle, Z., Rubenstein, J. L., and Sharpe, P. T. (1997). Role of *Dlx-1* and *Dlx-2* genes in patterning of the murine dentition. Development *124*, 4811-8.

Thompson, E. A. (1984). Information gain in joint linkage analysis. IMA J Math Appl Med Biol *1*, 31-49.

Vachon, G., Cohen, B., Pfeifle, C., McGuffin, M. E., Botas, J., and Cohen, S. M. (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*, 437-50.

van der Hoeven, F., Zákány, J., and Duboule, D. (1996). Gene transpositions in the *HoxD* complex reveal a hierarchy of regulatory controls. Cell *85*, 1025-35.

Weiss, K. M., Ruddle, F. H., and Bollekens, J. (1995). *Dlx* and other homeobox genes in the morphological development of the dentition. Connect Tissue Res *32*, 35-40.

Westerfield, M. (1995). The Zebrafish Book (Eugene: University of Oregon Press).

Williams, S. C., Altmann, C. R., Chow, R. L., Hemmati-Brivanlou, A., and Lang, R. A. (1998). A highly conserved lens transcriptional control element from the *Pax-6* gene. Mech Dev *73*, 225-9.

Wolberger, C., Pabo, C. O., Vershon, A. K., and Johnson, A. D. (1991a). Crystallization and preliminary X-ray diffraction studies of a MAT alpha 2-DNA complex. J Mol Biol *217*, 11-3.

Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., and Pabo, C. O. (1991b). Crystal structure of a MAT alpha 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. Cell *67*, 517-28.

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

Yu, G., Hsueh, Y., Zerucha, T., Nakafuku, M., Sheng, M., Ekker, M., and Rubenstein, J. L. R. (1999). A PDZ-Domain Protein That Is Expressed In The Embryonic Forebrain Co-Activates Transcription With The DLX2 Homeodomain Protein. Submitted to Cell March, 1999.

Zákány, J., and Duboule, D. (1996). Synpolydactyly in mice with a targeted deficiency in the *HoxD* complex. Nature *384*, 69-71.

Zappavigna, V., Renucci, A., Izpisua-Belmonte, J. C., Urier, G., Peschle, C., and Duboule, D. (1991). *HOX4* genes encode transcription factors with potential auto- and cross-regulatory capacities. EMBO J *10*, 4177-87.

Zerucha, T., Muller, J. P., Chartrand, N., and Ekker, M. (1997). Cross-interactions between two members of the *Dlx* family of homeobox- containing genes during zebrafish development. Biochem Cell Biol *75*, 613-22.

Zerucha, T., Stühmer, T., Park, B. K., Long, Q., Yu, G., Hatch, G., Gambarotta, A., Schultz, J., Rubenstein, J. L. R., and Ekker, M. (1999). A Highly Conserved Enhancer In The *Dlx5/Dlx6* Intergenic Region Is The Site Of Cross-Regulatory Interactions Between *Dlx* Genes In The Embryonic Forebrain. Submitted to Development March 1999.

Zhang, H., Hu, G., Wang, H., Sciavolino, P., Iler, N., Shen, M. M., and Abate-Shen, C. (1997). Heterodimerization of Msx and Dlx homeoproteins results in functional antagonism. Mol Cell Biol *17*, 2920-32.

Zhao, G. Q., Zhao, S., Zhou, X., Eberspaecher, H., Solursh, M., and de Crombrugghe, B. (1994). rDlx, a novel distal-less-like homeoprotein is expressed in developing cartilages and discrete neuronal tissues. Dev Biol *164*, 37-51.

Zhao, J. J., Lazzarini, R. A., and Pick, L. (1993). The mouse *Hox-1.3* gene is functionally equivalent to the *Drosophila Sex combs reduced* gene. Genes Dev *7*, 343-54.

Zhou, C., Shunji, C., Jinsheng, D., Junlin, L., Jori, G., and Milanesi, C. (1996). Apoptosis of mouse MS-2 fibrosarcoma cells induced by photodynamic therapy with Zn (II)-phthalocyanine. J Photochem Photobiol B *33*, 219-23.