uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Christine Weber

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Cellular and Molecular Medicine)

GRADE / DEGREE

Department of Cellular and Molecular Medicine

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Cdx-Hox Protein Interactions

TITRE DE LA THÈSE / TITLE OF THESIS

David Lohnes

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

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

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Ilona Skerjanc

Rashmi Kothary

Gary W. Slater

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

Cdx-Hox protein interactions

by

Christine Weber

This thesis is submitted as a partial fulfillment of the M.Sc. program in Cellular and Molecular Medicine

October 6, 2008

Department of Cellular and Molecular Medicine

Faculty of Medicine

University of Ottawa

Your file Votre référence
ISBN: 978-0-494-46504-2
Our file Notre référence
ISBN: 978-0-494-46504-2

Abstract

Cdx and *Hox* gene families both code for homeodomain-containing transcription factors and both are required for proper patterning of the anterior-posterior axis. Although Cdx is known to be upstream of certain *Hox* genes, preliminary data indicated Cdx1 interacts *in vitro* with Hoxd4 but not Hoxa9, and that this interaction is localized to the C-terminal region of both proteins. GST-pulldown assays with *in vitro* radiolabelled proteins or transfected cell lysates were used to investigate the extent of the Cdx-Hox interactions. The interaction was narrowed down to specific residues in the first half of the homeodomain, but an inhibitory N-terminus can abrogate interaction *in vitro*. A larger group of Hox proteins interact with GST-Cdx when they are expressed in a cellular environment. This could mean that the interaction requires other cofactors, such as Hox cofactors Pbx and Meis that also interact with GST-Cdx1 *in vitro*. This study is the first example of Cdx-Hox protein interactions and suggests that the Cdx-Hox complex could be involved in the regulation of downstream Hox target genes.

Table of Contents

List of Figures

List of Abbreviations

AP – anterior-posterior

ARE – autoregulatory element

AVE – anterior visceral endoderm

BMP – bone morphogenetic protein

C – cervical vertebra

Ca^{2+} – calcium

CYP26 – P-450 cytochrome oxidase

DBD – DNA binding domain

E – embryonic

ERK – elk-related tyrosine kinase

Exd – extradenticle

FGF – fibroblast growth factor

FGFR – fibroblast growth factor receptor

GFP – green fluorescent protein

GST – glutathione S transferase

HRP – horseradish peroxidase

JNK – c-jun kinase

LEF – lymphoid enhancer factor

LRP – low density lipoprotein-receptor related protein

MAPK – mitogen-activated protein kinase

PCP – planar cell polarity

PKC – protein kinase c

RA – retinoic acid

RALDH – retinaldehyde dehydrogenase

RAR – retinoic acid receptor

RARE – retinoic acid response element

RXR – retinoid x receptor

T – thoracic vertebra

TALE – three amino acid loop extension

TCF – T-cell factor

TGFβ – transforming growth factor beta

VAD – vitamin A deficient

VE – visceral endoderm

ZHX – zinc finger and homeobox

Acknowledgments

I would like to thank my supervisor, Dr. David Lohnes, for guiding and supporting my research for the past three and a half years. I owe to him my appreciation for the scientific research process and all of my laboratory skills. Thanks also to my fellow students, namely Donna, Colin, Jen, Stephanie, Jenny, and Beta, who all taught me as I taught them. Thanks to Joanne for her quiet wisdom, and Charlotte, James, Vicki, Catherine, and Sabrina for keeping the laboratory running smoothly.

I would also like to thank Dr. Kothary and Dr. Ekker for serving on my advisory committee, and Dr. Lee for his invaluable advice.

Thanks to my sister, Alison, and nephew, Callum, for being so supportive during the writing process. Special thanks to my close friend Sarah, who showed me that it can be done.

I would like to also thank all of my friends and family who supported me throughout my degree. Thank you to my parents, George and Sally, who always let me follow my own path and have never wavered in their unending support of everything I do.

Chapter 1

Introduction

Gastrulation is a major event in embryogenesis that results in the generation of the three germ layers (ectoderm, mesoderm, and endoderm), lays out the embryonic body plan along the AP axis, and initiates axis elongation (Iimura and Pourquie 2007). Once each germ layer is specified, it is typically patterned along the anterior-posterior (AP) axis to give rise to specialized derivatives.

In the mouse, the realization of the embryonic body plan is a result of events that occur before a morphologically overt body axis is apparent. In the mouse, the embryo resembles a cylindrical cup before gastrulation. Axis induction initiates by signals emanating from the visceral endoderm (VE), a layer of cells that underlies the epiblast and is later displaced to become yolk sac endoderm. The VE initially located at the distal end of the egg cylinder becomes thickened at embryonic day 5.5 (E5.5) and subsequently migrates to become anterior visceral endoderm (AVE) demarcating the future anterior end of the embryo (Rivera-Perez et al. 2003). The opposite side of the cylinder will become the posterior end of the embryo and is the site of gastrulation. Gastrulation begins at E6.5 by the ingression of cells through the primitive streak, a structure which initially forms at the posterior edge of the epiblast and subsequently expands towards the distal tip of the egg cylinder before regressing posteriorly. The epiblast cells that delaminate and transverse the primitive streak go on to become mesoderm and endoderm, while the ectoderm is generated by epiblast cells that do not ingress, thus creating the three germ layers of the embryo. After the primitive streak regresses, the generation of posterior structures is carried out by the tail bud.

In amniotes, the embryo develops in an anterior to posterior fashion, with anterior structures formed first while more posterior structures are formed as cells progressively

migrate through the primitive streak or later divide in the tail bud to contribute to the three body layers. Cells must be patterned such that more anterior cells will form anterior structures while the more posterior cells form more posterior structures. The Nieuwkoop model of activation-transformation (1952) hypothesizes that ectoderm is first "activated" by inducing signals to an anterior-only neural fate (Nieuwkoop 1952). These neural cells respond to a second "transforming" signal that caudalizes the cells to progressively more posterior fates. These transforming signals arise from the underlying mesoderm and are required for AP patterning of the neuraxis. It is thought that the same caudalizing molecules likely pattern mesoderm and endoderm in the same fashion. Subsequent generation of structures that differ along the AP axis is determined by so-called caudalizing pathways, which consist largely of Wnt, FGF and retinoic acid (RA)-mediated processes.

1.1 Wnt signaling

Wnts are a large family of extracellular signaling molecules that bind to members of the Frizzled and low density lipoprotein-receptor related protein (LRP) transmembrane receptor families to regulate gene expression by a variety of mechanisms.

The canonical Wnt pathway involves the cellular stabilization of β-catenin. In the absence of Wnt ligand, cytoplasmic β-catenin is associated with a degradation complex targeting it for ubiquitination. The presence of Wnt causes the dissociation of the degradation complex, resulting in cytoplasmic accumulation of β-catenin. β-catenin then translocates to the nucleus where it binds with lymphoid enhancer factor/T-cell factor (LEF/TCF) transcription factors to effect regulation of downstream targets (Figure 1.1a)

(as reviewed in (Logan and Nusse 2004)). LEF/TCFs bind to DNA via their HMG domains, and remain in a transcriptionally repressive state by association with repressors such as Groucho until β-catenin binds (Brantjes et al. 2001; Tutter et al. 2001). The canonical pathway is most often associated with the proliferation and differentiation abilities/capabilities of Wnt signaling.

Canonical Wnt signaling is involved in initiating AP patterning through the expression of *Wnt3* in the posterior visceral endoderm and subsequently in the posterior of the epiblast where it is required for primitive streak formation and proper orientation of the AP axis (Liu et al. 1999; Barrow et al. 2003). Wnt2b, -5a, -8a, and -11 are also expressed in a similar pattern to Wnt3 in the primitive streak (Kispert et al. 1996; Zakin et al. 1998; Yamaguchi et al. 1999b; Kemp et al. 2005).

Non-canonical Wnt signaling comprises any Wnt pathway that acts independently of β-catenin including convergent extension and the calcium pathway, steroid receptor binding, and atypical receptor tyrosine kinase pathways (Figure 1.1b) (Widelitz 2005). During axis initiation, the Wnt/calcium pathway antagonizes the canonical Wnt pathway through Wnt5a and Wnt11. The Wnt/planar cell polarity (PCP) pathway is involved in the axial elongation of the embryo by imparting positional information to a field of cells (Barrow 2006). The elongation of the AP axis is a result of convergent extension whereby cells intercalate along the medial-lateral axis during gastrulation. This results in cells extending posteriorly and lengthening the AP axis.

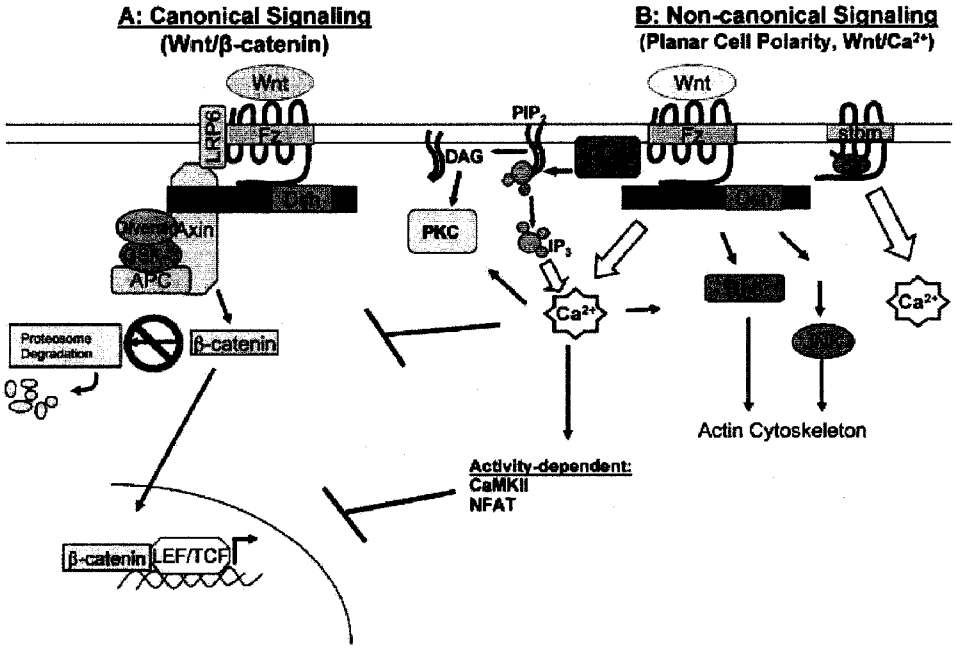

A: Canonical Signaling
(Wnt/β-catenin)
B: Non-canonical Signaling
(Planar Cell Polarity, Wnt/Ca2+)
Wnt
Wnt
LRP6
Fz
Fz
PIP2
DAG
PKC
IP3
Ca2+
Ca2+
Axin
APC
Proteosome Degradation
β-catenin
β-catenin
LEF/TCF
Actin Cytoskeleton
Activity-dependent:
CaMKII
NFAT

Figure 1.1 – Components of the Wnt Pathways. A) Canonical Wnt pathway. Upon Wnt binding to the Frizzled receptor, β-catenin translocates to the nucleus, and binds with LEF/TCF transcription factors to activate target gene expression. B) Non-canonical pathways act independently of β-catenin but still signal through Fz and Dsh. The Wnt/Ca^{2+} pathway involves an increase in intracellular Ca^{2+} and activation of PKC. The Wnt/PCP pathway signals through Rho and JNK to affect cytoskeletal conformation. Taken from Slusarski, D.C. and Pelegri, F. 2007. *Dev Biol* 307(1): 1-13.

1.2 Fibroblast growth factor signaling

Fibroblast growth factors (FGFs) are a large family of extracellular signaling molecules involved in many different cellular processes and are essential for proper development. They signal by activating FGF receptors (FGFRs), membrane bound receptor tyrosine kinases that dimerize, activate their tyrosine kinase domains and autophosphorylate upon FGF binding. There are a number of downstream pathways that can be activated by FGF signaling including MAPK, PI3K/Akt and PLC/Ca^{2+} pathways (Figure 1.2) (Bottcher and Niehrs 2005).

FGF signaling has an evolutionarily conserved role in driving the formation of mesoderm, cell movement through the primitive streak during gastrulation, and AP patterning (Yamaguchi et al. 1994; Deschamps et al. 1999; Sun et al. 1999). In mice, Fgf8 and Fgf4 are expressed in the epiblast and visceral endoderm at E5.75, and at E6.5 in the primitive streak of the embryo proper (Crossley and Martin 1995). In the absence of FGF signaling, mesoderm still forms, but has abnormal organization, not migrating away from the primitive streak (Yamaguchi et al. 1994; Sun et al. 1999). In *Xenopus*, overexpression of Fgf results in the anteriorization of mesoderm while inhibition of FGF signaling through a dominant-negative receptor results in defective posterior mesoderm, as evidenced through expression patterns of *Cdx* and *Hox* (discussed below)(Pownall et al. 1996; Haremaki et al. 2003).

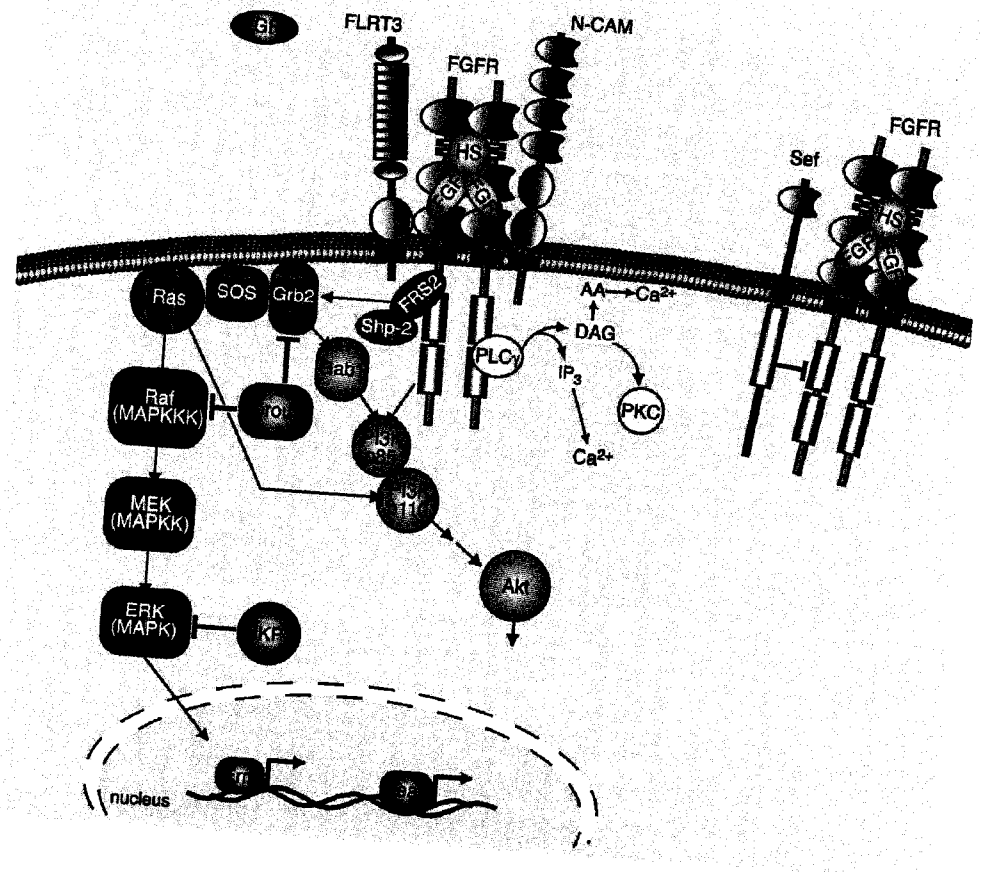

FLRT3
N-CAM
FGFR
FGFR
Sef
HS
Ras
SOS
Grb2
FRS2
Shp-2
PLCγ
AA→Ca2+
DAG
IP3
PKC
Ca2+
Raf
(MAPKKK)
MEK
(MAPKK)
ERK
(MAPK)
Akt
nucleus

Figure 1.2 – Components of the FGF Pathway. FGF ligands bind FGF receptors causing activation of a variety of intracellular signaling cascades. These include the MAPK (shown in blue), PI3K/Akt (shown in green), and PLC/Ca^{2+} (shown in yellow) pathways. Taken from Bottcher, R.T. and Niehrs, C. 2005. *Endocr Rev* 26(1): 63-77.

1.3 Retinoic Acid signaling

RA is the active form of vitamin A which must be obtained through diet as organisms do not produce it themselves. Vitamin A deficient (VAD) animals exhibit specific symptoms during adult life including weight loss and squamous metaplasia of epithelial tissues while the offspring of VAD mothers have specific congenital abnormalities, which demonstrate the importance of RA in both homeostasis and development (reviewed in Lohnes et al. 1995).

Vitamin A is transformed into active RA by retinal dehydrogenases (Raldhs) and is inactivated by members of the Cyp26 family of P450 cytochrome oxidases. The major members expressed during development are *Raldh2* and *Cyp26a1*. The opposing expression patterns of *Raldh2* and *Cyp26a1* create a gradient of RA along the AP axis of the developing embryo. *Raldh2* is first expressed at E7.5 in the primitive streak, and later during tail bud stages in the mesoderm. *Raldh2* mutants die at E10.5 with small somites, a shortened AP axis, hindbrain patterning defects and heart defects but can be partially rescued by administering RA *in utero* (Niederreither et al. 1997). *Cyp26a1* is normally expressed in the tail bud region and mutants exhibit defects concurrent with an excess of RA signaling, which include posterior homeotic transformations in the axial skeleton, where one vertebra takes on the morphological characteristics of another vertebrae (Abu-Abed et al. 2001; Sakai et al. 2001).

RA regulates gene expression by binding to an RA receptor (RARα, β and γ) that forms a heterodimer with a retinoid X receptor (RXRα, β, γ) partner. This heterodimer regulates transcription through a *cis*-acting regulatory sequence, the RA response element (RARE), which is present in the promoter region of target genes (Glover et al. 2006). In

the absence of RA, RAR-RXR heterodimers act as repressors through recruitment of histone deacetylase (HDAC) complexes that condense chromatin, making the DNA less accessible to both other transcription factors and the basal transcriptional machinery.

1.4 Patterning the AP axis – Somites

The information imparted from these pathways must be integrated and assimilated for proper development relative to axial level. As the body axis extends, paraxial mesoderm undergoes strictly timed segmentation on either side of the neural tube to generate paired epithelial structures called somites, a process known as somitogenesis (Figure 1.3). Somitogenesis occurs in a rostral to caudal fashion and is closely linked with axis extension.

As new somites are formed, the pool of mesodermal progenitors found in the tail bud extends posteriorly. The budding off of somites from the presomitic mesoderm occurs in a precise manner (every one hundred and twenty minutes in the mouse) at the determination front, a specific region along the AP axis that moves posterior as the axis extends. This rhythmic formation of somites is maintained by the segmentation clock which consists of a number of oscillating genes involved in the Wnt, Fgf, and Notch pathways (Dubrulle and Pourquie 2004).

Although they are morphologically identical upon formation, somites are already patterned with respect to their position along the AP axis, and go on to form structures with different morphologies dependent on their axial level. Somites subsequently differentiate into sclerotome, which give rise to the vertebrae, and dermamyotome, which form the musculature of the back and limbs, and the dermis of the trunk. There are five

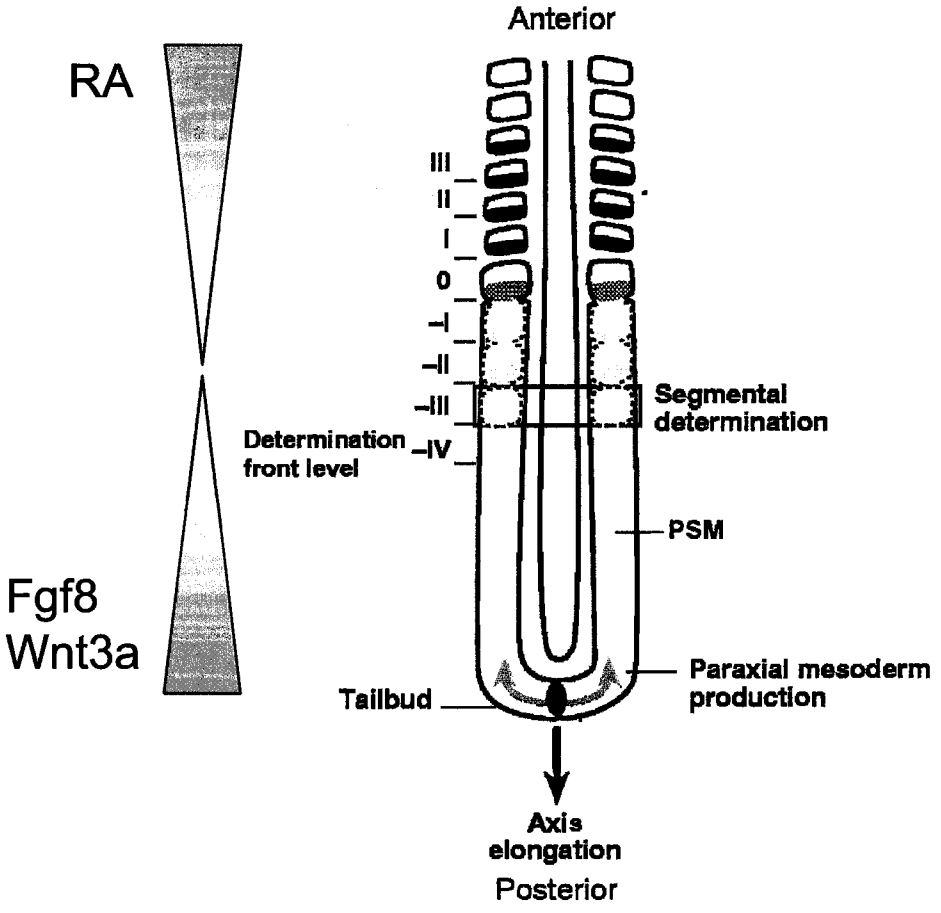

RA
Fgf8
Wnt3a
Anterior
III
II
I
0
–I
–II
–III
–IV
Determination front level
Segmental determination
PSM
Paraxial mesoderm production
Tailbud
Axis elongation
Posterior

Figure 1.3 – Gradients involved in driving AP patterning. Opposing gradients of RA (blue) and FGF8 and Wnt3a (pink) along the AP axis are involved in specifying the determination front. This is where presomitic mesoderm (PSM) becomes segmented into somites (see text for details). Adapted from Dubrulle, J. and Pourquie, O. 2004. *Development* 131(23): 5783-5793.

different types of vertebrae based on morphology. The cervical vertebrae are the most rostral vertebrae, followed by the thoracic vertebrae which have ribs. Following those are the lumbar, sacral, and caudal vertebrae. Each vertebral type has distinct characteristics, as do some individual vertebrae, as will be discussed below.

The differences in vertebral morphology along the AP axis are a result of the proper combination of transcription factors, which assimilate information from a variety of signals, resulting in differentiation of cells in the developing body through the regulation of target genes (Conlon 1995; Deschamps and van Nes 2005). *Hox* genes are a major target of the signaling pathways associated with caudalization, and the proper combination of *Hox* gene products is essential for appropriate AP patterning in the somites and vertebrae.

1.5 *Homeobox* genes

The *homeobox* gene family is a large family of genes that encode transcription factors involved in many processes including patterning the developing embryo along its axes. The homeobox is a highly conserved DNA sequence of 180 base pairs that codes for a 60 amino acid helix-turn-helix motif known as the homeodomain. The homeodomain consists of an N-terminal arm and three α-helices, the third of which is the site of specific DNA binding. Helices I and II lie parallel to each other and across from the third helix, which is also known as the recognition helix, and is the most highly conserved region of the homeodomain. The third helix binds the major groove of DNA while the N-terminal arm contacts specific residues of the minor groove (Gehring et al.

1994). The loop between helices I and II also contacts the DNA but is not involved in conveying specificity of binding.

There are a number of different subclasses of *homeobox* genes, the largest of which is the *Hox* class. *Hox* genes are unique in that they are present as clusters of genes in the genome and the relative location of a gene in the cluster predicts where it will be expressed along the AP axis of the developing embryo. At the protein level, Hox genes exhibit very high sequence similarity at the homeodomain. The 'extended Hox' or ParaHox class comprises genes (including the Cdx, Gsx, and Xlox families) that encode proteins with highly similar homeodomains to the *Hox* genes but are not present in the aforementioned clusters. The other homeobox containing gene classes are NK (which can also be clustered in the genome), Paired, LIM, POU, and atypical homeobox genes. These classes have diminishing degrees of conservation with the Hox homeobox; they may also have other conserved regions such as the paired, LIM, and POU domains that aide in DNA binding or protein-protein interactions. The atypical class contains divergent members of the homeodomain family. One subclass of the atypical homeodomains is the TALE (three amino acid loop extension) group, which is characterized by three additional residues in the loop between helix I and II, and whose members include the major Hox cofactors, as discussed below.

1.5.1 Evolution of homeobox clusters

Phylogenetic studies have shown that *homeobox* genes are evolutionarily conserved (Krumlauf 1992; Manak and Scott 1994). Indeed, the NK, ParaHox, and Hox clusters have been suggested to be derived from one common ancestral complex, the

Protohox cluster (Brooke et al. 1998). Hox clusters are an excellent model for studying both evolutionary and developmental biology. The evolution of the Hox cluster mirrors the evolution of animals, such that simpler organisms have fewer and simpler Hox gene clusters than complex organisms.

Cnidarians (such as jellyfish and corals) have two *Hox* genes, representing an anterior and a posterior class type; early bilaterians, such as acoel flatworms, have four *Hox* genes (Cook et al. 2004). *Drosophila* has eight *Hox* genes in a split cluster, while *Amphioxus*, a non-vertebrate chordate, has fourteen *Hox* genes. Among vertebrates, genome duplications resulted in multiple *Hox* clusters: four in humans and mice, seven in zebrafish, and up to fourteen clusters in tetraploid salmonoid fish (Moghadam et al. 2005).

Just as duplication events resulted in a more complex *Hox* cluster, duplication of the ProtoHox cluster resulted in the evolution of the 'extended Hox' class of homeobox genes. Prior to the expansion of the Hox cluster, the ProtoHox cluster duplicated, most likely in *cis* (the duplicate remains beside the original) to produce primordial ParaHox and Hox clusters. Extended Hox members are mostly scattered in the genome, although there are some that are still linked to the Hox clusters (Brooke et al. 1998).

1.6 *Hox* genes

Hox genes were first identified in *Drosophila* and are evolutionarily conserved among all animals investigated (Pearson et al. 2005). They are named based on the homeotic transformations that result due to the loss or gain of a particular homeobox gene which typically results in one segment assuming the morphological characteristics of

another body segment, an outcome particularly striking in *Drosophila* due to its obvious segments. For example, loss of *Antennapedia* results in ectopic antennae in the place of the second pair of legs while excess *Antennapedia* expression results in legs developing on the head in the place of antennae (Postlethwait and Schneiderman 1969). The segmented features of the vertebrate body plan are not as evident as in *Drosophila.* However, examples of segmentation can be seen along the AP axis in transient structures such as rhombomeres in the hindbrain and somites in the paraxial mesoderm, and later permanent structures that develop from them: for example, cranial nerves from the rhombomeres and vertebrae from the somites; changes in *Hox* expression result in AP patterning defects that affect most obviously the vertebrae.

A unique property of *Hox* genes is their property of collinear expression. This refers to the finding that more 3' genes of a given *Hox* cluster are expressed at more anterior boundaries and pattern more anterior structures. The genes located more 5' have more posterior anterior boundaries and are required for the patterning of more posterior structures (Figure 1.4).

In *Drosophila*, there is one split complex known as the HOM-C complex. In mammals, as a result of a series of duplications of the *Protohox* complex, there are four mammalian *Hox* clusters (HoxA through D) found on separate chromosomes. The genes in each cluster are separated into paralog groups (numbered 1 through 13, from 3' to 5') based on homeodomain sequence similarity and physical relation within a given cluster (Mcginnis and Krumlauf, 1992; Deschamps and van Nes, 2005). Not all paralog groups are represented in each cluster, and in humans and mice there is a total of 39 *Hox* genes. These genes are expressed sequentially along both the anterior-posterior (A-P) body axis

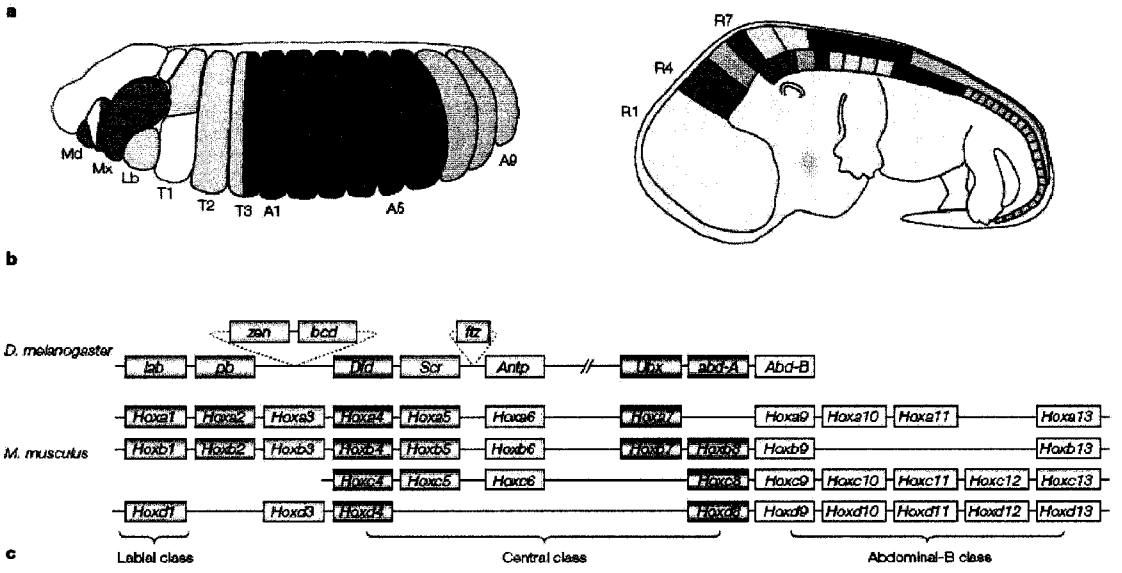

a
Md
Mx
Lb
T1
T2
T3
A1
A5
A9
R7
R4
R1
b
D. melanogaster
zen
bcd
fiz
lab
pb
Dfd
Scr
Antp
Ubx
abd-A
Abd-B
M. musculus
Hoxa1
Hoxa2
Hoxa3
Hoxa4
Hoxa5
Hoxa6
Hoxa7
Hoxa9
Hoxa10
Hoxa11
Hoxa13
Hoxb1
Hoxb2
Hoxb3
Hoxb4
Hoxb5
Hoxb6
Hoxb7
Hoxb8
Hoxb9
Hoxb13
Hoxc4
Hoxc5
Hoxc6
Hoxc8
Hoxc9
Hoxc10
Hoxc11
Hoxc12
Hoxc13
Hoxd1
Hoxd3
Hoxd4
Hoxd8
Hoxd9
Hoxd10
Hoxd11
Hoxd12
Hoxd13
c
Labial class
Central class
Abdominal-B class

Figure 1.4 – Conserved *Hox* expression and genomic organization. A) Representation of Drosophila (left) and mouse (right) embryos and the pattern of AP *Hox* expression during development. Md, mandibular; Mx, maxillary; Lb, labial; T, thoracic; A, abdominal; R, rhombomere. B) Schema of Hox clusters of both *Drosophila* and mouse. Note the genes are coloured to differentiate family members as well as represent where each group is expressed in the embryo proper in A). Adapted from Pearson, J.C., Lemons, D., and McGinnis, W. 2005. *Nat Rev Genet* 6(12): 893-904.

(Figure 1.4) (Deschamps and van Nes, 2005). Loss or gain of *hox* gene expression typically results in homeotic transformation of the morphology of the vertebral column, leading to the proposal of a "Hox code" (Kessel and Gruss, 1991; Mcginnis and Krumlauf, 1992), where the complement of *Hox* genes expressed at a given AP level is an indicator of the morphology of structures that will later form from the transient rhombomeres and somites.

Due to duplication of whole clusters, *Hox* paralog group members are more closely related than genes in the same cluster. With overlapping expression domains, loss of one paralog group member does not necessarily result in a phenotype, but additional loss of other group members can result in more severe phenotypes, indicative of functional redundancy.

As a result of the sequential activation of *Hox* genes along each cluster, cells in the posterior embryo express a larger number of *Hox* genes compared to the anterior cells. The most posteriorly expressed *Hox* gene is responsible for imparting patterning information, despite continued expression of more 3' genes, a mechanism known as posterior prevalence (Duboule 1991; Duboule and Morata 1994). Posterior prevalence is exemplified in murine mutants where one or a few body segments are affected, with the impacted region corresponding to the domain where the mutated *Hox* member is the most posterior *Hox* gene expressed (Kmita and Duboule 2003).

As mentioned above, *Hox* gene products are essential for patterning the somites/vertebrae, and shifts in *Hox* expression patterns result in shifts in the morphology of vertebrae. Thus, the diversity of vertebral columns among species can occur while still following the same basic patterning, as a result of shifts in *Hox* gene expression that do

not affect the entire axis. All vertebrates have the same vertebral regions along the AP axis (cervical, thoracic, lumbar, sacral, and caudal), although the number of vertebrae that contribute to each region differs among species. For example, mice have seven cervical vertebrae while chickens have fourteen. *Hox6* genes have anterior boundaries between somites 10 and 11 in mice and 18 and 19 in chicken, which corresponds to the site of transition between the cervical and thoracic vertebrae (Burke et al. 1995; Wellik 2007). *Hox* genes, therefore, have different anterior boundaries in different species but the boundaries still correspond to the same morphological regions.

1.6.1 Initiation of *Hox* expression

Hox genes are differentially expressed along the AP axis in all three body layers; the neurectoderm, paraxial and lateral plate mesoderm, and the endoderm derived gut tube all show nested patterns of *Hox* gene expression. Precise spatial and temporal initiation of *Hox* genes is essential for proper patterning of the AP axis (Deschamps and van Nes, 2005).

Hox genes are first expressed during gastrulation, such that cells that undergo gastrulation first are regionalized to the anterior of the embryo and express more 3' *Hox* genes. The later cells will contribute to more posterior structures and will express more posterior *Hox* genes (Iimura and Pourquie 2006). Indeed, it has recently been shown that *Hox* expression is initiated in the epiblast adjacent to the primitive streak, before cells undergo gastrulation; depending on which *Hox* gene was overexpressed, cells would pass through the primitive streak at different times and go on to contribute to structures at different axial levels, indicating that *Hox* genes may also regulate when cells ingress

through the primitive streak (Iimura and Pourquie 2006). This implies that *Hox* genes do more than pattern the AP axis; they are also involved in determining where cells will reside along the AP axis. Thus, *Hox* expression patterns and the AP identity of cells are determined before overt morphological differences are evident and even before the formation of somites.

Numerous upstream signals are required for proper *Hox* expression, and the nature of the events necessary to initiate and maintain the exquisite pattern of *Hox* expression have been the subject of much research (Brock and Fisher 2005; Deschamps and van Nes 2005; Pearson et al. 2005).

There are a number of mechanisms involved in the initiation of *Hox* genes in a colinear fashion. The chromatin itself must be modified through histone modifications to allow for transcriptional machinery to access promoter regions. This remodeling occurs in a sequential manner such that the more 3' *Hox* genes are revealed first and are thus available for transcription before more 5' genes. Global enhancer sequences located just outside of a *Hox* cluster are not promoter specific, but are responsible for differentially regulating the expression of several *Hox* genes. The distance of genes from these sequences may be an important factor in the regulatory asymmetry of these enhancers (Kmita and Duboule 2003). For example, a regulatory element on the 5' end of the *Hoxd* cluster has been shown to drive cycling expression during segmentation (Zakany et al. 2001). Finally, there are promoter-specific regions that are targets of signaling pathways as well as particular transcription factors.

1.6.2 Regulation of *Hox* expression by signaling pathways

As *Hox* expression is initiated during gastrulation, and primitive streak formation and regression, it follows that the molecular processes involved therein also regulate *Hox* gene initiation. Wnt signaling regulates the formation of the primitive streak and could also contribute to the regulation of the anterior progression of *Hox* genes (Ikeya and Takada 2001; Forlani et al. 2003). FGF signaling that causes movement of mesodermal cells upon ingression through the primitive streak also affects *Hox* expression (Ciruna and Rossant 2001). Overexpression of *Fgf4* causes ectopic anterior induction of *Hoxa7*, while overexpression of a dominant negative FGF receptor causes reduced expression of *Hoxa7* in *Xenopus laevis* (Pownall et al. 1996). Excess RA results in vertebral homeotic transformations that are accompanied by shifts in *Hox* gene expression (Kessel and Gruss 1991; Conlon and Rossant 1992). Indeed, many 3' *Hox* genes contain RAREs and are regulated by RA (Niederreither et al. 2000; Oosterveen et al. 2003). *Hox* expression is also mediated by Hox proteins themselves, either by auto- or cross-regulation. For example, Hoxa1 drives expression of *Hoxb1* in the neural tube, while Hoxb1 can auto-regulate thereafter (Popperl 1995; Studer et al. 1998).

The numerous signaling pathways that impact on Hox expression do not always do so directly. In such cases, the integration of signals must be conveyed through other transcription factors which then drive *Hox* expression. One such family is the *Cdx* family of homeobox genes.

1.7 *Cdx* genes

Cdx genes code for homeodomain transcription factors, are evolutionarily conserved, and play similar roles in posterior specification and/or AP patterning in all species examined. In non-vertebrates such as *Drosophila melanogaster* and *C. elegans*, *caudal/pal-1* is essential for patterning the posterior embryo (Moreno and Morata 1999; Baugh et al. 2005). *Caudal* is also required for proper AP patterning in *Drosophila*. Loss of *caudal* results in a homeotic transformation of the most posterior body segment, the analia, causing it to resemble the adjacent segment, the male genitalia (Moreno and Morata 1999). Conversely, overexpression of *caudal* in anterior segments results in ectopic development of analia (Moreno and Morata 1999). In vertebrates, there are three *caudal* homologues: *Xcad-1*, *Xcad-2*, and *Xcad-3* in *Xenopus laevis* (Blumberg et al. 1991; Northrop and Kimelman 1994); *CdxA*, *CdxB*, and *CdxC* in chicken (Frumkin et al. 1991; Serrano et al. 1993; Marom et al. 1997); and *Cdx1*, *Cdx2*, and *Cdx4* in humans (Bonner et al. 1995; Horn and Ashworth 1995; Drummond et al. 1997; Mallo et al. 1997) and mice (Gamer and Wright 1993; Meyer and Gruss 1993; Beck et al. 1995).

1.7.1 Expression Patterns and null phenotypes

1.7.1.1 Cdx1

Cdx1 transcripts are first observed at embryonic day (E) 7.5 in the ectoderm and nascent mesoderm of the primitive streak, and reach an anterior expression boundary in the presumptive hindbrain by E7.75 (Meyer and Gruss 1993). These expression boundaries extend laterally in the posterior streak and by E8.5 the entire posterior embryo expresses *Cdx1* with expression levels decreasing rostral to the node. At this stage, *Cdx1*

is also expressed in the neural tube and somites. At E10.5, *Cdx1* is restricted to the dorsal somite, the dermamyotome. In addition, *Cdx1* is expressed in the proximal forelimb bud at E9.5, and at lower levels in the hindlimb bud at E10.5. *Cdx1* continues to be expressed in the primitive streak and tail bud, with expression extinguishing around E12 (Figure 1.5) (Meyer and Gruss 1993).

Cdx1 is also expressed in the intestinal epithelium beginning at E14.0 and continuing into adulthood (Duprey et al. 1988). This initial expression in the intestine correlates with a squamous to columnar transition of the intestinal epithelium and the formation of villi. In the adult, expression levels increase from the duodenum to the distal colon (James et al. 1994).

Although viable and fertile, *Cdx1* null mutants exhibit vertebral homeotic transformations affecting the cervical skeleton (Figure 1.6) (Subramanian et al. 1995). In the wild type mouse there are 7 cervical vertebrae (C1 – C7), most of which can be morphologically distinguished. The first cervical vertebra (C1) of *Cdx1* null mice is frequently anteriorized and fused to the basioccipital bone, while the second vertebra (C2) has broadened neural arches and an ectopic anterior arch of the atlas (AAA), features which are typical of C1. Consistent with anterior homeotic transformation of the cervical region, the third vertebra (C3) exhibits thickened neural arches resembling C2 (Subramanian et al. 1995). In the thoracic region, which is characterized by the presence of ribs, the second thoracic vertebra has a dorsal spinous process which has shifted to the third thoracic vertebra in *Cdx1* null mutants. In addition, some *Cdx1* null mice also have an additional rib or pair of ribs attached to the sternum (Subramanian et al. 1995).

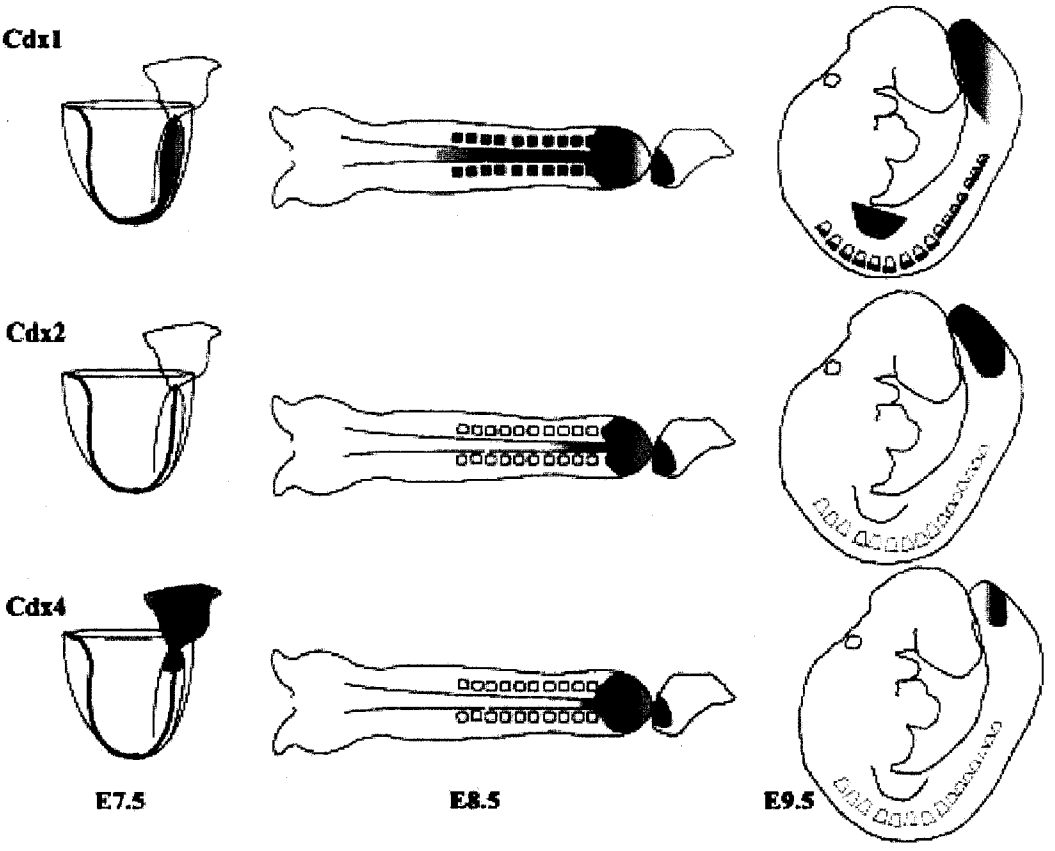

Cdx1
Cdx2
Cdx4
E7.5
E8.5
E9.5

Figure 1.5 – Expression patterns of mouse *Cdx* genes. Schema representing expression patterns from E7.5 to E9.5 in the embryo proper. Note the graded expression patterns such that *Cdx1* is expressed most anteriorly and *Cdx4* has the most posterior region of expression (see text for details). Taken from Lohnes, D. 2003. *Bioessays* 25(10): 971-980.

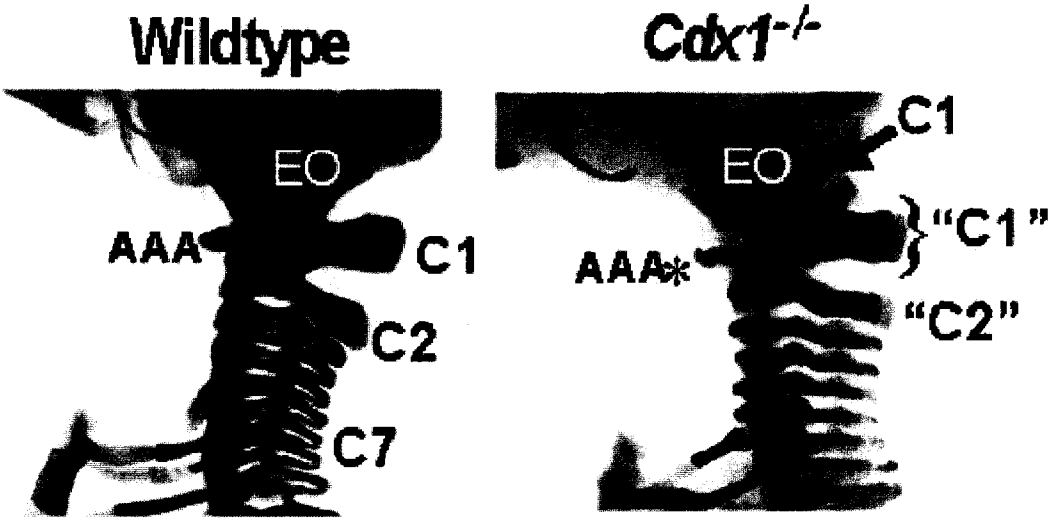

Wildtype
Cdx1-/-
EO
AAA
C1
C2
C7
EO
C1
"C1"
AAA*
"C2"

Figure 1.6 – Homeotic transformation of $Cdx1^{-/-}$ vertebrae. Lateral view of a homeotic axial transformation, where the first cervical vertebra, C1, is fused to the occipital bone and the second and third cervical vertebrae have taken on the characteristics of C1 and C2, respectively. EO, exoccipital; AAA, anterior arch of the atlas; C, cervical vertebra; *, ectopic AAA. Quotation marks indicate presumptive transformations. Adapted from Lohnes, D. 2003. *Bioessays* 25(10): 971-980.

A relationship between Cdx1 and regulation of *Hox* expression is evidenced by a posterior shift of one to two somites of the anterior boundary of expression of a number of *Hox* genes in *Cdx1* null mutants and by Cdx response elements in the promoter regions of a number of *Hox* promoters (Subramanian et al. 1995). Interestingly, there is no intestinal phenotype (Subramanian et al. 1995), the only site of adult *Cdx1* expression.

1.7.1.2 Cdx2

Cdx2 is first expressed in the trophectoderm at E3.5 and is critically required for proper extra-embryonic development as *Cdx2* null mutants die around E4.5 due to implantation defects (Chawengsaksophak et al. 1997). After implantation, expression continues in trophectoderm-derived ectodermal layer of the chorion and the spongy cytotrophoblasts of the placenta (Beck et al. 1995). In the embryo proper, *Cdx2* expression is first seen at E8.5 in all tissues of the primitive streak and at the base of the allantois (Figure 1.5), where it remains until E12.5. At this stage, *Cdx2* expression is confined to the gut endoderm and the posterior tip of the tail bud. Like *Cdx1, Cdx2* expression increases greatly in the embryo at the time of transition of the gut epithelium from stratified squamous epithelium to the characteristic columnar epithelium and the concomitant formation of the villi in the small intestine and crypts in the large intestine (James et al. 1994). In the adult, *Cdx2* is broadly expressed in the intestinal epithelium in a gradient increasing from duodenum to proximal colon, then decreasing from proximal colon to rectum as well as in the endocrine pancreas, which is also derived from the endoderm (James et al. 1994).

Cdx2 null mice are not viable due to peri-implantation lethality (Chawengsaksophak et al. 1997). *Cdx2* heterozygotes are viable and fertile, with over half exhibiting a shortened or kinked tail of variable severity (Chawengsaksophak et al. 1997). Like *Cdx1* mutants, *Cdx2* heterozygotes exhibit anterior vertebral homeoses and altered patterns of *Hox* gene expression. However, consistent with the later onset of *Cdx2* compared with *Cdx1*, these vertebral homeoses begin at the sixth cervical vertebra (C6) (Chawengsaksophak et al. 1997). For example, *Cdx2* heterozygotes display anterior tubercules, normally present on C6, on the C7. In addition, the spinous process normally found on the second thoracic vertebra (T2) is located on T3 in some *Cdx2* heterozygotes. There are also rib abnormalities, such that the eighth ribs attach to the sternum, either unilaterally or bilaterally (Chawengsaksophak et al. 1997).

Cdx2 heterozygotes also have an intestinal phenotype, with the majority of offspring developing lesions predominantly in the proximal colon, where intestinal *Cdx2* expression is highest (Beck et al. 1999; Chawengsaksophak et al. 1997). These areas cease to express *Cdx2*, due to extinction of expression from the remaining allele, as opposed to loss of heterozygosity (Chawengsaksophak et al. 1997). The finding that Cdx2 heterozygotes exhibit an intestinal phenotype while Cdx1 mutants do not is most likely due to the fact that the expression levels of *Cdx2*, compared to *Cdx1*, in the intestine is approximately tenfold at this stage of development (James et al. 1994).

Using aggregation of $Cdx2^{-/-}$ embryonic stem cells with wild type tetraploid cells from the morula stage, the implantation defect can be circumvented. The tetraploid cells will only contribute to extraembryonic structures, including the trophectoderm, which requires *Cdx2* expression, allowing the embryo proper, populated entirely by $Cdx2^{-/-}$

cells, to continue developing beyond implantation. However, mutant embryos only survive until E11.5, as the allantois fails to fuse with the chorion (Chawengsaksophak et al. 2004). These mutants are developmentally delayed compared to controls, and are caudally truncated. Somites are small and irregular posterior to somite 5, and there is little to no unsegmented paraxial mesoderm at the posterior tip of the embryo. However, the neural tube and endoderm do extend to the posterior tip of the embryo (Chawengsaksophak et al. 2004). Again, like *Cdx1* null mice, these *Cdx2* mutants show alterations in the expression levels of some *Hox* genes (e.g. *Hoxb1* and *Hoxd4*), and shifts in the anterior boundary of *Hoxb8* expression (Chawengsaksophak et al. 2004).

$Cdx1^{+/-}Cdx2^{+/-}$ mice exhibit a more severe phenotype than either heterozygote alone (van den Akker et al. 2002). There is a higher penetrance of anterior homeosis affecting the cervical and thoracic vertebrae, in addition to defects not previously seen in either mutant impacting the sacral vertebrae and ribs. The anterior defects in the anterior cervical vertebrae are not as penetrant as in $Cdx1^{-/-}$ mice, suggesting that *Cdx1* is the predominant Cdx member in patterning this region. This is in agreement with expression data (Meyer and Gruss 1993; Beck et al. 1995), however, $Cdx1^{-/-}Cdx2^{+/-}$ mutants exhibit a more severe phenotype than either $Cdx1^{-/-}$ or $Cdx1^{+/-}Cdx2^{+/-}$ mice along the AP axis, including the cervical region, indicating that there is a role for *Cdx2* in cervical vertebral patterning even though the Cdx2 heterozygotes do not display any phenotype at this level (van den Akker et al. 2002). In addition, these mutants exhibit an increased shortening of the AP axis, with some mutants having only a few caudal vertebrae posterior to the hindlimbs (van den Akker et al. 2002). The anterior boundary of *Hox* genes is affected in *Cdx1/Cdx2* compound mutants. The more posterior *Hox* genes (*Hoxb8*, *Hoxb9*) are

affected by the loss of both *Cdx1* and *Cdx2* while a shift in *Hoxd4* expression is seen with loss of *Cdx1* alone (van den Akker et al. 2002).

1.7.1.3 Cdx4

Cdx4 is localized to the X chromosome (Horn and Ashworth 1995) which results in only one functional allele due to X-inactivation in females. In contrast with the other *Cdx* genes, *Cdx4* is not expressed in the adult and to a lesser extent in the embryo. *Cdx4* is first expressed at mid to late gastrulation (E7.0 – 7.5) in the allantois and the posterior tip of the primitive streak (Gamer and Wright 1993). By E8.5, *Cdx4* is present in all three germ layers in the posterior embryo, with transcripts extending anteriorly in the paraxial mesoderm to just posterior of the most recently formed somite, and extending to the level of the second or third most recently condensed somite in the neural tube and in the hindgut epithelium. The maximum level of *Cdx4* expression is in the posterior tip of the embryo (Figure 1.5), with expression subsequently extinguished by E10.5 (Gamer and Wright 1993). In this regard, it is notable that Cdx members exhibit a nested expression pattern of *Cdx* genes during gastrulation with a posterior-high distribution.

Cdx4 null ($Cdx4^{-/0}$) mice are viable and healthy with no apparent abnormalities (van Nes et al. 2006). However, $Cdx1^{-/-}/Cdx4^{-/0}$ and $Cdx2^{+/-}/Cdx4^{-/0}$ double mutants exhibit AP patterning defects consistent with functional overlap between Cdx members (van Nes et al. 2006), as the skeletal defects seen in $Cdx1^{-/-}$ and $Cdx2^{+/-}$ mice are exacerbated by the additional loss of *Cdx4*. $Cdx2^{+/-}/Cdx4^{-/0}$ compound mutants also exhibit axial truncations more severe than $Cdx2^{+/-}$ offspring (van Nes et al. 2006), indicating Cdx4 is also involved in axial elongation. Moreover, like $Cdx2^{-/-}$ tetraploid

aggregation mutants (Chawengsaksophak et al. 2004), *Cdx2*$^{+/-}$/*Cdx4*$^{-/0}$ mice also fail to develop past E10.5 due to lack of chorio-allantoic fusion (van Nes et al. 2006).

Thus, compound mutants show that there are overlapping, redundant functions among *Cdx* gene products. Although the contribution of each *Cdx* gene is not equal due to differences in expression patterns and less *Cdx4* transcripts as a result of being localized to the X-chromosome, compound mutants show more severe phenotypes than single mutants (van den Akker et al. 2002; van Nes et al. 2006). In addition, there is no overt phenotype when *Cdx2* is knocked in to the *Cdx1* locus, indicating that Cdx2 can functionally replace Cdx1 if expressed in its place (J. Savory, our laboratory, unpublished observation). Thus, specific function such as the anterior-most axial expression of *Cdx1* and the extra-embryonic expression of *Cdx2* could be due solely to expression patterns and not differential protein function.

1.7.2 Regulation of *Cdx* expression

Analysis of the regulation of *Cdx1* shows that it is a target of several upstream signals (Lohnes 2003). RA directly regulates *Cdx1* via an atypical RARE in the proximal promoter, and it has been proposed that this is an important mechanism for RA-dependent regulation of mesodermal *hox* expression (Houle et al., 2000). Functional mutation of this RARE results in downregulation of *Cdx1* expression in the developing embryo with phenotypic impact, and is thus crucial to normal *Cdx1* function. Other elements are, however, likely involved, as *Cdx1* still responds to RA in the absence of this RARE (Houle et al. 2003).

Cdx1 is also regulated by Wnt signaling during gastrulation/tail bud stages as well as in the intestine (Lickert et al. 2000; Prinos et al. 2001). The *Cdx1* proximal promoter has two functional LEF/TCF binding sites which respond to Wnt signaling, and are critical for *Cdx1* expression and function *in vivo* (Pilon et al. 2007; Prinos et al. 2001). Wnt3a and RA synergize on the *Cdx1* proximal promoter, further implicating Cdx1 as an important modulator of posteriorizing signals (Prinos et al. 2001). This proximal promoter is sufficient for driving the early expression of *Cdx1* (Lickert and Kemler 2002), with maintenance of expression dependent on a positive autoregulatory loop (Prinos et al. 2001) which operates through physical interaction between Cdx1 and LEF1 (Beland et al. 2004).

Cdx2 binds to its own promoter at the TATA box, as well as at an AT-rich region which have both been suggested to be involved in autoregulation (Xu et al. 1999). Other tissue-specific *Cdx2* enhancer elements located distal to the proposed autoregulatory site have been described that direct mesodermal and neural expression patterns. These enhancer regions predict possible regulation by Wnt, FGF, Sox, and Cdx (Wang and Shashikant 2007).

Recent evidence shows *Cdx4* is also directly regulated by the Wnt pathway as *Cdx4* responds to Wnt3a *in vivo*. Also, there are LEF/TCF response elements in the *Cdx4* promoter that respond to LEF1 and β-catenin (Pilon et al. 2006).

In other species, *Cdx* genes are known to be regulated by RA, Wnt, and FGFs. Zebrafish *cdx1a* and *cdx4* have been shown to be regulated by Wnt as well as FGF signaling (Shimizu et al. 2005), while *Xenopus Xcad3* is regulated by RA, Wnt, and FGF (Pownall et al. 1996; Haremaki et al. 2003; Shiotsugu et al. 2004). FGF also regulates

cdxA and cdxB in the chick (Bel-Vialar et al. 2002). FGF regulation has not been directly proven, but may play a role in murine *Cdx* regulation.

Taken together, Cdx members integrate a number of caudalization signals and contribute to the patterning of the posterior embryo.

1.7.3 Regulation of *Hox* by Cdx

Among other targets, Cdx proteins directly regulate *Hox* genes, a function that is essential for proper AP *Hox* expression patterns, as evidenced by the vertebral homeosis and concomitant shifts of *Hox* expression in various *Cdx* mutants (Gaunt et al. 1988; Subramanian et al. 1995; Pownall et al. 1996; Charite et al. 1998; Isaacs et al. 1998; Gaunt et al. 2004). Indeed, the defects seen in *Cdx* mutants are highly similar to *Hox* loss-of-function at relevant vertebral elements. For example, *Hoxd4* is expressed in the paraxial mesoderm up to the fifth somite, and *Hoxd4* null mice exhibit incomplete penetrance of a C2 to C1 transformation, fusion of C1 to the basioccipital bone, and malformations of the neural arches of C1, C2 and C3 (Horan et al. 1995a; Folberg et al. 1997); these defects are highly reminiscent of the *Cdx1* null phenotype (Figure 1.6) (Subramanian et al. 1995).

In addition to the above example, a number of *Hox* genes harbour Cdx binding motifs (TTTATA/G) (Margalit et al. 1993; Suh et al. 1994) in their enhancer regions, including *Hoxa5*, *Hoxa7*, *Hoxb8*, and *Hoxc8*, some of which have been shown to be functionally relevant *in vivo* (Knittel et al. 1995; Subramanian et al. 1995; Charite et al. 1998; Tabaries et al. 2005).

Multimerization studies of *Hox* enhancer regions has led to the hypothesis that *Hox* genes respond to different levels of Cdx, such that more 3' *Hox* genes are more sensitive to Cdx and thus respond to a lower amount of Cdx (Charite et al. 1998; Gaunt et al. 2004). For example, when the *Hoxa7* enhancer is multimerized, the expression boundary of the *Hoxa7/lacZ* reporter gene is anteriorized due to additional Cdx binding elements (Gaunt et al. 2004). Anteriorization of *Hox* expression could also occur due to an increase in Cdx expression (Epstein et al. 1997). Taken together, Cdx proteins activate a number of *Hox* genes with different anterior boundaries; the caudal to rostral gradient of Cdx causes the more sensitive 3' *Hox* genes to be expressed more anteriorly than the less sensitive 5' *Hox* genes (Gaunt et al. 2004).

1.8 Protein-protein (homeodomain-homeodomain) interactions

Dimerization presumably allows a small number of transcription factors to regulate complex gene expression patterns by binding to more varied regulatory sequences than monomer binding alone could accomplish. Indeed, physical interactions between homeodomain proteins have been widely described. For example, homodimerization has been demonstrated for Oct1 (Poellinger et al. 1989), Cdx2 (Suh et al. 1994), even-skipped (Hirsch and Aggarwal 1995), Mix.1 (Mead et al. 1996), and Pit1 (Jacobson et al. 1997).

All three members of the ubiquitously expressed zinc finger and homeobox (ZHX) family, ZHX1, ZHX2, and ZHX3, contain two zinc finger domains and five homeodomains. These transcriptional repressors both homo- and hetero-dimerize via

regions within the first homeodomain of both proteins (Kawata et al. 2003a; Kawata et al. 2003b; Yamada et al. 2003).

Members of the Msx and Dlx homeodomain protein families also interact through their homeodomains. However, in this case, interaction and DNA binding appear to be mutually exclusive, as both interactions are mediated by the same homeodomain residues (Zhang et al. 1996; Zhang et al. 1997). Msx proteins are transcriptional repressors, while Dlx proteins are transcriptional activators, and heterodimerization between these members counteracts their inherent transcriptional function, thus providing another level of transcriptional regulation in tissues where they are co-expressed (Zhang et al. 1997). Msx1 similarly binds with Lhx2, a LIM-type homeoprotein, through the homeodomains of both proteins (Bendall et al. 1998). As with the Msx-Dlx interaction, the Msx1-Lhx2 interaction occurs to the detriment of DNA binding. Therefore, Msx can affect gene expression not at the level of direct repression via DNA binding, but instead by sequestering transcriptional activators through protein-protein interactions (Bendall et al. 1998).

Taken together, the transcriptional abilities, both activating and repressing, of homeodomain proteins in general can be enhanced through homodimerization as well as heterodimerization between homeoproteins.

1.8.1 Hox binding partners

Although Hox proteins can bind DNA as monomers (Hoey and Levine 1988), they do so weakly and there is no consensus DNA binding site for Hox monomers even though the homeodomain is highly conserved. The specificity of Hox DNA binding

comes in part via its association with partners such as Pbx and Meis family members. *Pbx* and *Meis* code for atypical homeodomain proteins and have a three amino acid loop extension (TALE) between the first and second homeodomain helices.

Pbx is related to *Drosophila* Extradenticle (Exd), which was originally identified as a mutant which caused homeotic transformations without affecting *Hox* expression. This indicates that Exd works in parallel with, and not upstream of, Hox (Peifer and Wieschaus 1990; Rauskolb et al. 1995). In vertebrates, there are four *Pbx* family members (*Pbx1-4*) that functionally overlap (Chang et al. 1995). *Pbx* members have different expression patterns in the developing embryo, and individual *Pbx* null mutants typically display phenotypes related to where they are the prevalent (or only) *Pbx* member expressed (Selleri et al. 2001; Kim et al. 2002; Rhee et al. 2004; Selleri et al. 2004). For example, *Pbx1* mutants have severe organ defects as well as cervical vertebral malformations resembling the *Hoxa3*, *Hoxd3*, *Hoxa4*, and *Hoxa6* mutants (Condie and Capecchi 1994; Kostic and Capecchi 1994; Selleri et al. 2001).

In vertebrates, as in *Drosophila*, Pbx proteins function as Hox cofactors and recognize a conserved Hox/Pbx binding site, TGATTNAT, which is recognized by the heterodimer (Popperl 1995). Pbx interacts with Hox through its TALE loop and binds to the hexapeptide, or YPWM, motif found N-terminal to the homeodomain in *Hox* paralog groups 1-10 (Chang 1995; Phelan and Featherstone 1997). Although Pbx can physically bind to Hox in the absence of the hexapeptide motif (Galant et al. 2002; Merabet et al. 2003), some Hox functions still require Pbx (Remacle et al. 2004).

The MEIS class of TALE homeoproteins includes Meis and Prep in vertebrates, both of which are related to *Drosophila* Homothorax (Hth). Like Pbx, Meis members

also regulate Hox activity and are required for Pbx/Exd function in the absence of DNA (Chang et al. 1997; Berthelsen et al. 1998). Meis/Prep proteins interact directly with Pbx both within the DNA-bound Hox-containing complex (Mann and Affolter 1998; Jacobs et al. 1999; Ryoo and Mann 1999) and also affect nuclear localization and stability of Pbx proteins themselves (Berthelsen et al. 1999; Jaw et al. 2000; Longobardi and Blasi 2003). Originally found to interact with Hox through the hexapeptide motif, Meis/Prep proteins can also interact with members from paralog groups 9-13, which do not harbor a hexapeptide motif. The functionality of these interactions is, as yet, unknown (Shen et al. 1997; Williams et al. 2005a). Hox/Pbx/Meis complexes involve Hox/Pbx binding to one binding site, and Meis binding a separate site nearby (Chang et al. 1997).

1.8.2 Cdx-protein interactions

Cdx1 and Cdx2 have been shown to interact with other transcription factors, including homeodomain proteins, as well as with constituents of the basal transcriptional machinery. Cdx2 has been shown to interact with the homeoproteins Brn-4, Hnf1a, and Pax6 to drive expression of intestine and pancreas specific genes (Ritz-Laser et al. 1999; Mitchelmore et al. 2000; Wang et al. 2006). Cdx2 and Brn-4 interact to activate *proglucagon* (Wang et al. 2006), while Cdx2 and Hnf1a synergize on the lactase promoter (Mitchelmore et al. 2000). There are most likely relevant interactions on additional promoters, as both of these transcription factors regulate a large number of intestine specific genes (Mitchelmore et al. 2000). In the case of Brn-4 and Hnf1a, the interaction domain includes the homeodomain regions of both proteins (Mitchelmore et al. 2000; Wang et al. 2006).

Cdx2 and Pax6 synergize on the *glucagon* promoter, in association with the co-activator p300 (Hussain and Habener 1999). Of note, Cdx2 is able to enhance the physical interaction between Pax6 and p300 on the promoter, in the absence of binding to DNA itself although the interface between Cdx2 and Pax6 is as yet uncharacterized (Hussain and Habener 1999). Cdx1 also interacts with TATA-binding protein (TBP) via its homeodomain on the *glucose-6-phosphatase* promoter (Gautier-Stein et al. 2003). In addition to TBP, Cdx1 interacts with TAF7, TAF12, TAF15, and Med7, members of two of the major transcriptional activation complexes, by co-immunoprecipitation (Calon et al. 2007). Recently, Pbx1 has also been implicated as a co-factor of Cdx2 for regulating *proglucagon* in pancreatic A cells (Liu et al. 2006). This interaction is mediated through a penta-peptide motif of Cdx2, reminiscent of Pbx-Hox association (Chang et al. 1995; Phelan and Featherstone 1997; Liu et al. 2006). Cdx1 is also essential for its own expression (Prinos et al. 2001). Although the *Cdx1* promoter does not harbour a Cdx1 binding site, Cdx1 contributes to its own expression through a physical interaction with LEF1 (Beland et al. 2004). This interaction depends on sequences in the Cdx1 homeodomain and the B-box of LEF1 (Beland et al. 2004). Thus, Cdx members interact with a number of cofactors that are required for proper targeting of downstream genes.

1.9 Rationale

Protein-protein interactions are a way to increase specificity of DNA binding and activation of target genes. Hox proteins are highly conserved in the DNA binding homeodomain region, necessitating cofactors to increase specificity. The *Cdx* family exhibits similar expression domains as the more posterior *Hox* genes. As well, both

families are involved in AP patterning, both can directly regulate *Hox* expression, and both families can bind cofactors (some of which are identical) in the presence or absence of DNA. Therefore, it was postulated that Cdx might be another Hox cofactor, affecting expression of downstream targets. Initial *in vitro* pulldown analysis showed that Cdx1 binds Hoxd4, the C-terminal homeodomain region of both proteins being sufficient for association (Beland, 2004).

As Hox proteins are known to auto- and cross-regulate, and Cdx1 and Cdx2 have been shown to regulate specific *Hox* genes (Knittel et al. 1995; Subramanian et al. 1995; Charite et al. 1998; Tabaries et al. 2005), the *Hox* promoters themselves are potential targets for Cdx-Hox interactions.

1.10 Hypothesis and Objectives

I hypothesize that Cdx and a subset of Hox proteins interact through their respective homeodomains. This hypothesis will be addressed by (i) testing representative Hox proteins for interaction with Cdx1, (ii) using sequence alignment to elucidate the Hox residues responsible for conferring binding, and (iii) testing potential Hox promoter regions to convey biological significance to these findings.

CHAPTER 2

Materials and Methods

2.1 Expression vectors

Hox sequences were amplified by PCR and subcloned into pCR®2.1 according to the manufacturer's instructions (Invitrogen TA Cloning® Kit). To introduce a FLAG-tag, the homeodomain fragments were excised using restriction sites introduced in the primers used for initial amplification, subcloned into pCEP4flagΔCla·RV (modified from Invitrogen), and then into p514 (a derivative of pSG5 (Stratagene)(Green et al. 1988)) which allows for both expression in both mammalian cells in culture and *in vitro* from a T7 RNA polymerase binding site present in the vector.

Hox N-terminal and C-terminal homeodomain sequences were amplified from genomic DNA by PCR, while full length Hox cDNAs were generated by reverse-transcriptase PCR (RT-PCR) with the exception of FLAG-tagged Hoxd4, which was generously provided by Mark Featherstone. For RT-PCR, total RNA was isolated from (E9.5) mouse embryos using TRIzol® reagent (Invitrogen), and cDNA was synthesized by reverse transcription using oligo$(dT)_{12\text{-}18}$ (Invitrogen) and Moloney murine leukemia virus (M-MLV) reverse transcriptase (Invitrogen). Products were then subcloned as above to generate FLAG-tagged expression vectors. PCR primers for C-terminal *Hox* sequences are as follows:

Hoxa1Cterm [5'-TAAAAAGCTTCCCAACGCAGTGCGC-3' (f), 5'-ATTTGGATCCTCAGTGGGAGGTAGTC-3' (r)],

Hoxb1Cterm [5'- TCGTAAGCTTCCCGGCGGTCTCCGCACAAAC-3' (f), 5'-ACAAGGATCCTCAAGAGGTGATGGAACTGGG-3' (r)],

Hoxa4Cterm [5'-CTATAAGCTTCCTAAGCGCTCTCGAAC-3' (f), 5'-CATTGGATCCTTATATGGAGGAGGG-3' (r)],

Hoxd4Cterm [5'-TAGCAAGCTTCCCAAGCGCTCCCGGACGGCCTACACC-3' (f), 5'-TGCAGCTAGCCTATAAGGTCGTCAGGTCCGTATGG-3' (r)], Hoxa5Cterm [5'-TTCTAAGCTTGGCAAAAGGGCCCGG-3' (f), 5'-CATTGGATCCTCAGGGGCGGAAAGC-3' (r)], Hoxa6Cterm [5'-TATTAAGCTTGGGCGCAGAGGCCGC-3' (f), 5'-AGATGGATCCCTACTCGCCCGCTTTG-3' (r)], Hoxc8Cterm [5'-TCTTAAGCTTCGGCGCAGCGGTCGACAAAC-3' (f), 5'-CAAGTGGATCCGAAACTTCAAGGGAGTTGCTGGGG-3' (r)], Hoxd8Cterm [5'-TTCTAAGCTTAGACGGAGAGGAAGAC-3' (f), 5'-TACTGGATCCTTAATTTGTGGGGCAGC-3' (r)], Hoxa9Cterm [5'-GTGCAAGCTTACTCGGAAGAAGCGATGCC-3' (f), 5'-CGACCTCGAGTCACTCGTCTTTT-3' (r)], Hoxd10Cterm [5'-TCTTAAGCTTGGCAGAAAGAAGAGG-3' (f), 5'-TATCGGATCCCTAAGAAAAGGTGAG-3' (r)], Hoxd12Cterm [5'-TTCTAAGCTTGCCCGCAAGAAGAGG-3' (f), 5'-ACTAGGATCCCTAATAGAGGGCCAG-3' (r)], and Hoxd13Cterm [5'-TTCTAAGCTTGGAAGGAAGAAAAGG-3' (f), 5'-CTACGGATCCTCAGGAGACAGTGTC-3' (r)]. Hoxd4Nterm primers used are [5'-TAGCAAGCTTATGGCCATGAGTTCGTATATGGTG-3' (f), and 5'-TGCAGCTAGCCTACTCCCCGCCGGTGTAGTTGG-3' (r)].

Amplification of full length *Hox* sequences used the reverse primer of C-terminal sequences and forward primers as follows:

Hoxb1 [5'-ATTAGCGGCCGCGACTATAATAGGATGAGTTCC-3' (f)],

Hoxd4 [5'-TACTAAGCTTAGTTCGTATATGGTGAACTCTAAGTAC-3' (f)], Hoxa6 [5'-TCACGCGGCCGCAGTTCCTATTTTGTGAATCCC-3' (f)], Hoxc8 [5'-GTTACAAGCTTAGCTCCTACTTCGTCAACCCC-3' (f)], and Hoxa9 [5'-GTCTAAGCTTGCCACCACCGGG-3' (f)].

The glutathione S-transferase (GST)-Cdx1 fusion proteins have been described previously (Beland et al. 2004). FLAG-Meis and Pbx1a were generous gifts from Mark Featherstone.

2.1.1 Fluorescent fusion proteins

Hoxd4 was amplified by PCR using primers [5'-TACAGGTACCAGTTCGTATATGGTGAACTCTAAGTACGTG-3' (f)] and [5'-TGCAGGATCCCTATAAGGTCGTCAGGTCCGTATGG-3' (r)] then subcloned into dsRed (Invitrogen) to generate dsRed-Hoxd4. Cdx1 was amplified by PCR using primers [5'- CTAGAATTCTATGTACGTGGGCTATGTGCTG-3' (f)] and [5'-CTAGGATCCCTAGGGTAGAAACTCCTCCTTG-3' (r)] then subcloned into EGFP (Invitrogen) to generate EGFP-Cdx1.

2.1.2 Hoxb1 mutations

Primers corresponding to altered *Hoxb1* sequences were used to generate *Hoxb1* mutant vectors using the QuikChange site-directed mutagenesis kit (Stratagene). All mutants were verified by sequencing. The primers used to generate the G→R point mutant (residue 201) were [5'-CTGGGCGCTCCCGGCCGGCTCCGCACAAACTTC-3' (f)] and [5'-GAAGTTTGTGCGGAGCCGGCCGGGAGCGCCCAG-3' (r)] (mutated

sequences underlined). The primers used to generate the double K,A→R,R mutant (residues 222 and 227) were [5'-GAATTTCATTTCAACAGGTACCTGAGCCGTCGCCGGAGGGTGGAGATC-3' (f)] and [5'-GATCTCCACCCTCCGGCGACGGCTCAGGTACCTGTTGAAATGAAATTC-3' (r)]. In order to create a G,K,A→R,R,R triple mutant, the double mutant (R,R) construct was subjected to site-directed mutagenesis using the G→R mutant primers.

2.1.3 Hox-Hox Fusion Proteins

Fusion proteins between Hox members were created by amplifying relevant cDNA sequences in a two-step PCR process, whereby the first PCR generates individual fragments of each gene and the second PCR results in a fusion of these two fragments, which can then be used to translate a protein that has components of two different genes.

Generation of Hoxa6d4hom used the full length Hoxa6 forward primer (Hoxa6 (f), above) and [5'-GCGGCCTCTGCGCCCCTCCCCGCCGGTGTAGTTGG-3' (r)] to generate the N-terminal fragment, while the C-terminal fragment was generated using [5'-GTGTATGGGAGTCACCCCAAGCGCTCCCGGACGGC-3' (f)] and the Hoxd4Cterm reverse primer (Hoxd4Cterm (r), above). Hoxd4a6hom was generated using the full length Hoxd4 forward primer (Hoxd4 (f)) [5'-TACTAAGCTTAGTTCGTATATGGTGAACTCTAAGTAC-3' (f)] and [5'-CCGGGAGCGCTTGGGGTGACTCCCATACACGGCAC-3' (r)] to derive the N-terminal fragment, and [5'-TACACCGGCGGGGAGGGGCGCAGAGGCCGCCAGAC-

3' (f)] with the Hoxa6Cterm reverse primer (Hoxa6Cterm (r)) to generate the C-terminal fragment.

Similarly, Hoxb1d4hom used the full length Hoxb1 forward primer (Hoxb1 (f)) and [5'-GCGGAGACCGCCGGGCTCCCCGCCGGTGTAGTTGG-3' (r)] to generate the N-terminal fragment, and [5'-CTGGGACTGGGCGCTCCCAAGCGCTCCCGGACGGT-3' (f)] and Hoxd4Cterm reverse primer (Hoxd4Cterm (r)) to generate the C-terminal fragment. Hoxd4b1hom was derived using the full length Hoxd4 forward primer (Hoxd4 (f)) and [5'-CCGGGAGCGCTTGGGAGCGCCCAGTCCCAGCTCGG-3' (r)] to generate the N-terminal fragment, and [5'-TACACCGGCGGGGAGCCCGGCGGTCTCCGCACAAA-3' (f)] and the Hoxb1Cterm reverse primer (Hoxb1Cterm (r)) to generate the C-terminal fragment.

Hoxb1d4hII used the full length Hoxb1 forward primer (Hoxb1 (f)) and [5'-TGAGCGATTTCAATCCGACGGGCACGGCTCAGGTA-3' (r)] to generate the N-terminal fragment, and [5'-TACCTGAGCCGTGCCCGTCGGATTGAAATCGCTCA-3' (f)] and Hoxd4Cterm reverse primer (Hoxd4Cterm (r)) to generate the C-terminal fragment. Hoxd4b1hII uses the full length Hoxd4 forward primer (Hoxd4 (f)) and [5'-GCGGCGATCTCCACCCTCCGGCGCCTGGTCAGATA-3' (r)] to generate the N-terminal fragment, and [5'-TATCTGACCAGGCGCCGGAGGGTGGAGATCGCCGC-3' (f)] and the Hoxb1Cterm reverse primer (Hoxb1Cterm (r)) to generate the C-terminal fragment.

These initial fragment sets were employed as substrates in a second PCR step to generate the fusion sequences using primer sets (sequences above) as follows:

Hoxa6 (f) and Hoxd4Cterm (r) for Hoxa6d4hom,

Hoxd4 (f) and Hoxa6Cterm (r) for Hoxd4a6hom,

Hoxb1 (f) and Hoxd4Cterm (r) for Hoxb1d4hom and Hoxb1d4hII, and

Hoxd4 (f) and Hoxb1Cterm (r) for Hoxd4b1hom and Hoxd4b1hII.

2.2 Plasmid Preparation

Plasmids were transformed into Subcloning Efficiency™ DH5α™ Competent Cells according to manufacturer's instructions (Invitrogen). Subsequently, the plasmid DNA was isolated using QIAGEN® Plasmid Purification kits, following the guidelines for a Maxi preparation.

2.3 GST fusion protein purification

GST fusion proteins were purified as previously described (Beland et al. 2004). Briefly, BL-21 bacteria were transformed with the base GST plasmid or GST-Cdx1 fusion construct. Starter cultures were grown for 3 hours, treated with 0.5 mM IPTG (Bioshop), and grown for another 3 hours. Cells were then pelleted and resuspended in PBS with 1 mM DTT (EMD) and protease inhibitors (1 μg/mL aprotinin (Roche), 1 μg/mL leupeptin (Sigma), 1 μg/mL pepstatin A (Sigma), 1 mM PMSF (Sigma)), denoted PBS+i. Cells were lysed by sonication using a Branson Sonifier 450, 1% Triton-X (Sigma) was added to the lysate and incubated on ice for 15 minutes. Lysate was centrifuged again at 13000 rpm and the supernatant incubated with BD BaculoGold glutathione-agarose beads (BD Biosciences) prepared as per the manufacturers instructions for 1-2 hours at 4°C. The beads were then washed 3 times with PBS+i. For

quantification, an aliquot of the beads were resolved by SDS-PAGE and proteins revealed by Coomassie stain.

2.4 Tissue culture and transient transfection assays

Cos7 cells were grown in Dulbecco's modified Eagle's medium (D-MEM) supplemented with 10% heat-inactivated fetal bovine serum (FBS). P19 embryocarcinoma cells were grown in α-MEM supplemented with 2.5% FBS and 7.5% heat-inactivated donor calf serum (DCS).

Cos7 transfections were carried out using Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions, using 6μL Lipofectamine 2000 per 10 cm plate. For Cos7 transfections, 5 μg of experimental plasmid and 1 μg of a GFP expression vector (to monitor transfection efficiency) were transfected onto confluent 10 cm plates. Cell lysates from Cos7 cells were harvested 24 h later by scraping in ice-cold PBS and resuspended in 300 μL lysis buffer (20 mM Tris [pH 8.0], 25 mM NaCl, 1.5 mM $MgCl_2$, 1 mM EGTA, 1% Triton X-100, 10% glycerol, 1 mM DTT, and protease inhibitors) for 30 min.

P19 cells were passaged into 6-well plates at approximately 400,000 cells per well for reporter assays. Cells were transfected using the calcium phosphate precipitation method (Gibco), using 0.5 - 1 μg reporter vector, 0.02-0.5 μg expression vectors, and 0.5 μg of a GFP expression vector (to monitor transfection efficiency). Cells were harvested 48 hours post-treatment and the lysates processed and analyzed for luciferase activity using the Promega Luciferase Assay System, according to the manufacturer's instructions.

For analysis of fluorescent-tagged proteins, Cos7 or P19 cells were passaged into 6-well plates containing glass coverslips (Fisher) coated with 0.1% gelatin for 1 hour. The cells were transiently transfected with 0.5 μg expression vectors using the calcium phosphate precipitation method (Gibco) and fixed 48 hours post-treatment with 4% paraformaldehyde (PFA). Coverslips were then washed in PBS and mounted onto glass slides (Fisher) using Prolong® Gold antifade reagent with DAPI (Invitrogen) and subcellular distribution assessed using an Axiophot Microscope.

2.5 Protein interaction assays

2.5.1 *In vitro* pulldown assays

In vitro protein-protein interaction assays were conducted as previously described (Beland et al. 2004). Briefly, ^{35}S-methionine labeled Hox proteins were generated *in vitro* using the TNT Quick Coupled Trancription/Translation system (Promega). 5 μL of ^{35}S-labelled protein was then incubated with 5 μg of GST fusion protein (affixed to glutiothione-agarose) in 750 μL TNENi (50 mM Tris, 100 mM NaCl, 0.1 % Igepal, 5 mM EDTA, 1 mM DTT, protease inhibitors) overnight at 4°C. The beads were then washed 3 times, proteins resolved by SDS-PAGE, and Hox proteins revealed by autoradiography. For pulldown assays involving Pbs or Meis, the glutathione agarose beads were washed in a higher stringency buffer (200 μM NaCl).

2.5.2 Pulldown assays from cell lysates

Hox proteins were produced by transient transfection in Cos7 cells (explained above). Fifty μL of cell lysate was incubated with 5 μg of GST fusion protein affixed to

glutiothione agarose beads in 750 μL TNENi overnight at 4°C. The beads were then washed 3 times and associated proteins resolved by SDS-PAGE. Protein was then transferred to Immobilon-P membrane (Sigma) for one hour in transfer buffer (39 mM glycine, 48 mM Tris Base, 20% methanol) for analysis by Western blotting. Membranes were then blocked for one hour in blocking solution (2% blocking powder (Amersham) in Tris-buffered saline with 0.1% Tween-20 (TBS-T)), rinsed with TBS-T, and incubated overnight at 4°C with α-Flag (Sigma) at a 1:5000 dilution in blocking solution. Membranes were then washed 3 x 10 minutes in TBS-T, incubated with secondary antibody goat anti-mouse HRP (Santa Cruz) at 1:50000 dilution in blocking solution for one hour, and washed again 3 x 10 minutes in TBS-T. Protein was then revealed using the advance ECL kit (Amersham) according to manufacturer's instructions and developed using Kodak X-Omat film (Kodak).

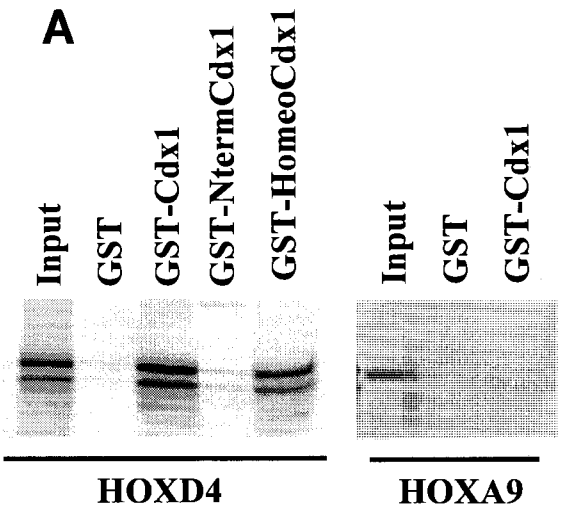
CHAPTER 3

Results

As Cdx1 and Hoxd4 mutants have similar morphological abnormalities at the cervical vertebral level, it was hypothesized that Cdx1 could be involved in the regulation of targets of the *Hox4* group (including Hox genes themselves) by association with Hox proteins. Indeed, *in vitro* translated Hoxd4 interacted with GST-Cdx1, but not with GST alone (Figure 3.1a). In contrast, initial findings indicated that Hoxa9 did not interact with GST-Cdx1 *in vitro*, suggesting some specificity among the Hox proteins with respect to interaction with Cdx1.

To identify the region of Cdx1 necessary for interaction with Hoxd4, GST fusion proteins containing either the Cdx1 homeodomain (residues 163-214) or the N-terminus (amino acids 1-160) were used. Hoxd4 from transiently transfected Cos7 cell lysate interacted specifically with the homeodomain of Cdx1, but not with N-terminal sequences (Figure 3.1a). This is consistent with the interaction between Cdx1 and LEF1, which also requires the Cdx homeodomain (Beland et al. 2004), and represents the first observation of direct interactions between members of these two transcription factor families (Subramanian et al. 1995; Pownall et al. 1996; Charite et al. 1998; Isaacs et al. 1998; Gaunt et al. 2004).

The conserved regions among Hox proteins are the YPWM motif and the homeodomain, both of which are known to be involved in protein-protein interactions (Goutte and Johnson 1994; Chang et al. 1995). In order to determine the Hox residues which are required for interaction with Cdx1, Hoxd4 was separated into two fragments, and N-terminal region (residues 1-151), which included the YPWM motif, and the homeodomain-containing C-terminal (152-250). Only the C-terminal fragment interacted with GST-Cdx1 *in vitro* (Figure 3.1b). As the sequence C-terminal to the homeodomain


A
Input
GST
GST-Cdx1
GST-NtermCdx1
GST-HomeoCdx1
Input
GST
GST-Cdx1
HOXD4
HOXA9

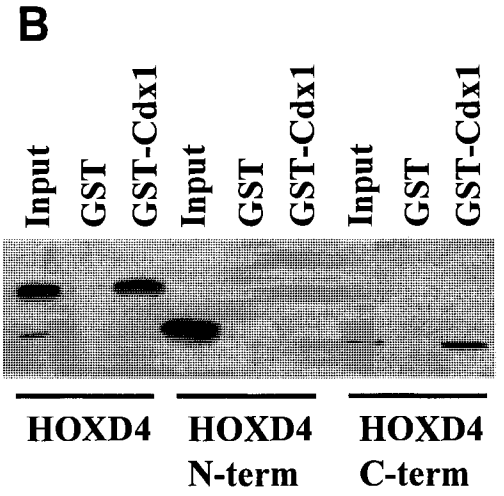

B
Input
GST
GST-Cdx1
Input
GST
GST-Cdx1
Input
GST
GST-Cdx1
HOXD4
HOXD4
N-term
HOXD4
C-term

Figure 3.1 – Identification of interacting regions of Hoxd4 and Cdx1. A) Autoradiograph of GST-Cdx1 pulldowns with ^{35}S-labelled full length Hox proteins. Hoxd4 interacts strongly with Cdx1 while Hoxa9 does not interact at all. Input represents 20% of ^{35}S-labelled protein used in pulldowns. B) Western blot of GST-Cdx1 pulldowns with protein extracts from transfected Cos7 cells. Hoxd4 full length and C-terminal fragment interact with GST-Cdx1 while Hoxd4 N-terminal does not. Input represents 5% of the lysate used (Béland, 2004).

is not conserved between Cdx members, and both Cdx1 and Cdx2 interact with Hoxd4 the site of interaction between Hoxd4 and Cdx1 is most likely mediated through their respective homeodomains.

3.1 Subcellular localization of Cdx1 and Hoxd4

To confirm that Cdx1 and Hoxd4 are capable of physically interacting in cells, the subcellular localization of these proteins was examined. Due to the lack of specific Hox antibodies, fluorescently tagged *Cdx1* and *Hoxd4* expression vectors were transfected into either P19 or Cos7 cells to visualize their location within the cell. Both transcription factors co-localized to the nucleus, and can be seen together in nucleoli in both cell types (Figure 3.2), suggesting that Cdx1 and Hoxd4 are physically present in the same subcellular location of the cell.

3.2 Characterization of physical interactions between a subset of Hox proteins and Cdx *in vitro*

The interactions between Cdx1 and Hoxd4, and non-interaction with Hoxa9, is indicative of Cdx1 interacting with a subset of Hox proteins. To further investigate this, Hox proteins from a range of paralog groups were assayed for the interaction with Cdx1. As the area of interaction with Hoxd4 was confined to the C-terminus, and this region is found in one exon, initial assays employed expression vectors harboring this domain which were generated by PCR amplification of genomic DNA. To this end, relevant sequences from representative members of *Hox* paralog groups 1, 4, 5, 6, 8, 9, 10, 12, and 13 were amplified, subcloned to acquire an epitope (flag) tag and T7 and SV40 promoters


A
Merge
B
Merge

Figure 3.2 – Subcellular localization of Hoxd4 and Cdx1. Direct fluorescence shows transiently transfected dsRed-Hoxd4 and EGFP-Cdx1 colocalize to the nucleus in both Cos7 (A) and p19 (B) cells.

to drive expression *in vitro* and in cell culture, respectively. These constructs were then expressed by *in vitro* translation in the presence of ^{35}S-methionine and used in pulldown assays with GST-Cdx1. The majority (Hoxa4Cterm, Hoxd4Cterm, Hoxa5Cterm, Hoxa6Cterm, Hoxd8Cterm, Hoxd10Cterm, Hoxd12Cterm, Hoxd13Cterm – Cterm signifies C-terminus) of such Hox sequences interacted strongly with Cdx1, while Hoxa1Cterm and Hoxa9Cterm were weak interactors and Hoxb1Cterm and Hoxc8Cterm did not interact detectably (Figure 3.3). Thus, interaction with Cdx1 is restricted to a subset of Hox homeodomain fragments.

In order to assess whether full length Hox proteins exhibited the same binding profiles as the cognate homeodomains, full length *Hox* cDNA sequences were amplified by RT-PCR and subcloned as per the C-terminal regions. Full length Hox proteins were then derived by *in vitro* translation in the presence of ^{35}S-methionine and assessed for interaction with GST-Cdx1 *in vitro*. This analysis revealed that, as with the homeodomain sequences, full length Hoxd4 interacted with Cdx1, while Hoxb1, Hoxa6, Hoxc8 and Hoxa9 proteins did not (Figure 3.4). This indicates that Hox proteins that are weak, or non-interactors as C-terminus fragments do not interact with Cdx1 as full length proteins. This observation also further confirms that the site of Hox interaction with Cdx1 is restricted to the C-terminus.

3.2.1 N-terminal Hox sequences can inhibit interaction with Cdx1

Interestingly, *in vitro* binding assays also revealed a class of Hox proteins which interacted with Cdx1 as homeodomains, but not as full length proteins. For example, the isolated C-terminus of Hoxa6 interacted with Cdx1, but this interaction no longer

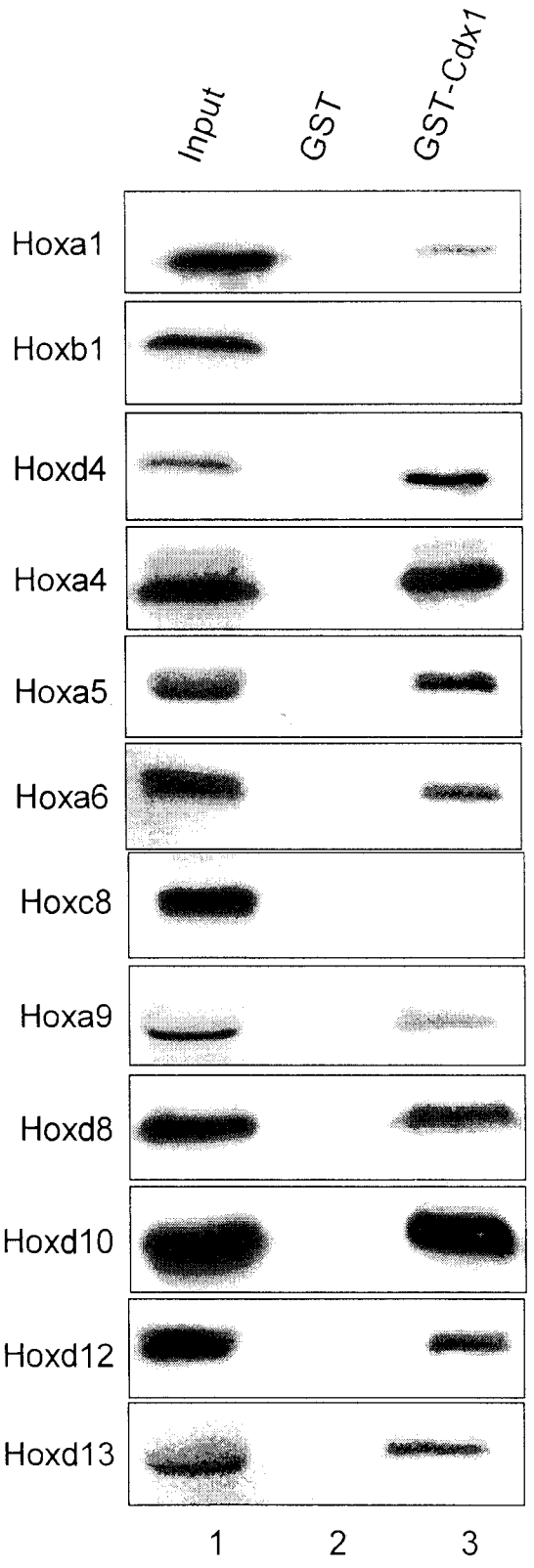

Input
GST
GST-Cdx1
Hoxa1
Hoxb1
Hoxd4
Hoxa4
Hoxa5
Hoxa6
Hoxc8
Hoxa9
Hoxd8
Hoxd10
Hoxd12
Hoxd13
1
2
3

Figure 3.3 – Interaction of Hox homeodomains with GST-Cdx1. ^{35}S-labelled Hox homeodomains made with Promega T7 quick coupled transcription/translation kit were incubated with glutathione agarose beads bound to either GST or GST-Cdx1 fusion protein. The beads were then washed and resolved on a SDS-PAGE gel. Results were exposed by autoradiography. Input represents 20% of ^{35}S-labelled protein used in pulldowns.

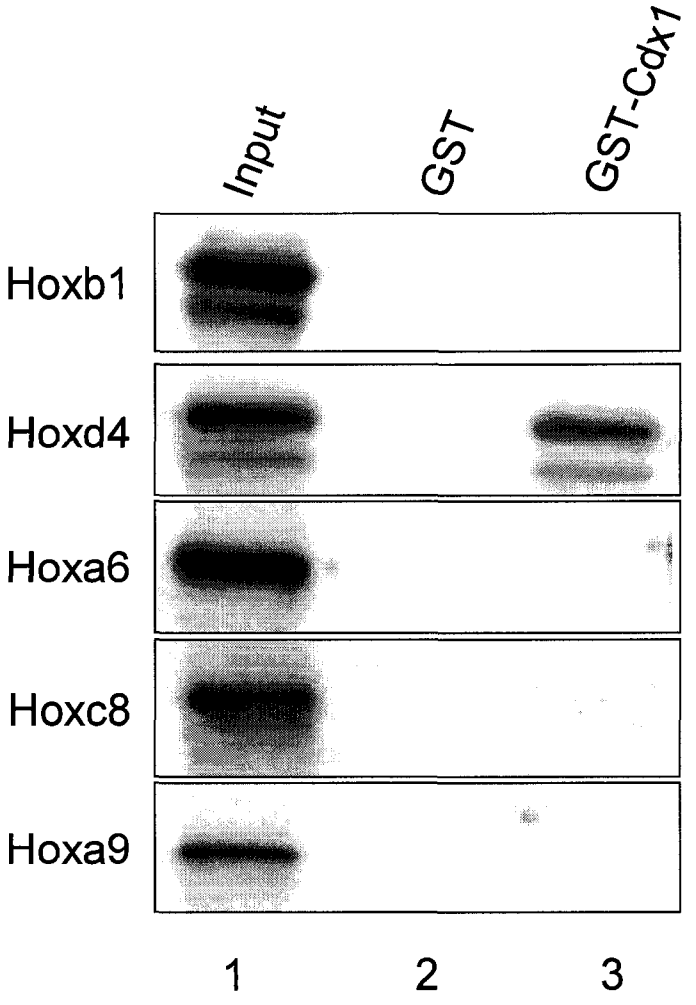

Input
GST
GST-Cdx1
Hoxb1
Hoxd4
Hoxa6
Hoxc8
Hoxa9
1
2
3

Figure 3.4 – Interaction of full length Hox proteins with GST-Cdx1. Autoradiograph of GST-Cdx1 pulldowns with ^{35}S-labelled full length Hox proteins. Hoxd4 interacts strongly with Cdx1, while Hoxb1, Hoxa6, Hoxc8, and Hoxa9 do not interact. Input represents 20% of ^{35}S-labelled protein used in pulldowns.

occurred with the full length protein (Figures 3.3,3.4). This lead to the hypothesis that N-terminal Hox sequences may inhibit the interaction. To investigate this, chimeric proteins between Hoxd4 (which interacts with Cdx1 as both a homeodomain fragment and full length protein) and Hoxa6 (which interacts only as a homeodomain protein) were constructed. Hoxd4 N-terminal sequence (residues 1-151) fused with the Hoxa6 C-terminal sequence from the beginning of the homeodomain to the end of the protein (residues 154-232) is denoted as Hoxd4a6hom, while Hoxa6d4hom is the converse construct. *In vitro* pulldown assays revealed that Hoxd4a6hom interacted with GST-Cdx1 while Hoxa6d4hom did not (Figure 3.5), even though the Hoxd4 C-terminal region is able to interact with Cdx1 both in isolation and in the context of the full length Hoxd4 sequence (Figure 3.4). Thus, this observation suggests that the N-terminal region of Hoxa6 (and potentially other Hox proteins) inhibits the ability of some Hox homeodomains to interact with Cdx1.

3.3 Characterization of homeodomain sequences permissive for Cdx interaction

The above observations demonstrate that permissive Hox N-terminal sequences and specific homeodomain characteristics are both required for interaction of full length Hox proteins with Cdx1. Both full length Hoxd4 as well as its homeodomain interact with Cdx1, while Hoxb1 does not interact in either form. Therefore, to further understand the differences between interacting and non-interacting Hox homeodomains, fusion proteins between Hoxd4 and Hoxb1 were generated. The Hoxd4b1hom construct is the Hoxd4 N-terminus residues (1-151) fused to the Hoxb1 homeodomain (199-297) while the Hoxb1d4hom is Hoxb1 N-terminus (1-198) fused to Hoxd4 C-terminus (152-250).

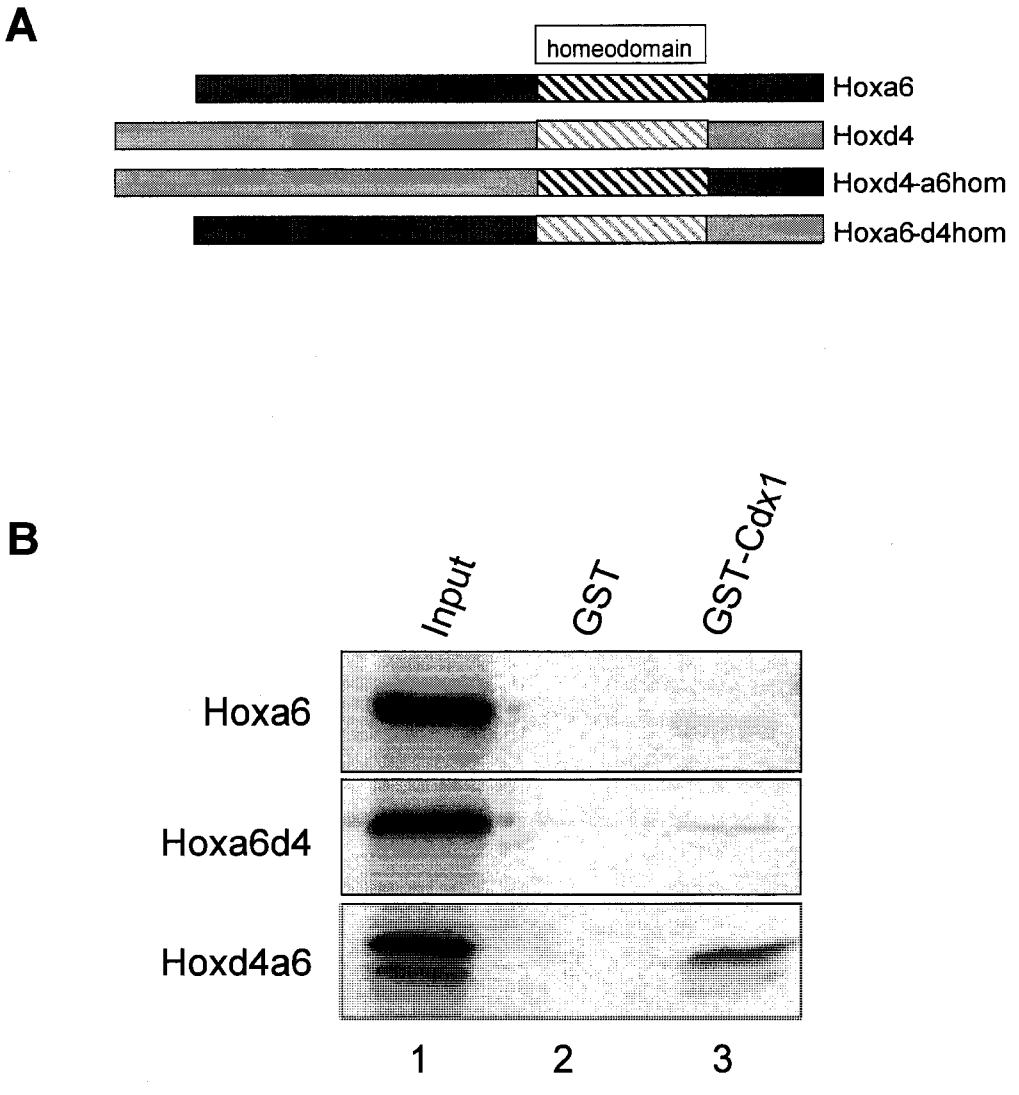

A
homeodomain
Hoxa6
Hoxd4
Hoxd4-a6hom
Hoxa6-d4hom
B
Input
GST
GST-Cdx1
Hoxa6
Hoxa6d4
Hoxd4a6
1
2
3

Figure 3.5 – Identification of inhibitory N-terminal regions in Hoxa6. N-terminal regions of Hoxa6 are inhibitory to the Cdx-Hox interaction. Autoradiograph of ^{35}S labelled Hox proteins incubated with GST or GST-Cdx1. Hoxd4 and Hoxd4b1 interact with GST-Cdx1 while Hoxb1 and Hoxb1d4 do not. 20% ^{35}S labelled protein as input.

Pulldown assays revealed that the Hoxb1d4hom chimeric protein interacted with GST-Cdx1, while the Hoxd4b1hom fusion protein did not. This is in agreement with the C-terminal data, where the Hoxd4 C-terminus can interact with GST-Cdx1, whereas the Hoxb1 C-terminus cannot (Figure 3.6). Of note, these data also demonstrate that the Hoxb1 N-terminus does not have inhibitory functions that abrogated interaction of full length Hoxa6 with GST-Cdx1. Rather, in this case, the lack of interaction is likely due to the inability of the Hoxb1 homeodomain itself to bind Cdx1.

To further delineate sequences required for interaction with Cdx1, a second set of chimeric proteins was generated, fusing Hox sequences between the first and second helices of the homeodomain. As the third helix is the site of Hox DNA binding, and is highly conserved, it is likely that the selective interactions seen in a subset of Hox proteins is due to helix 3 sequences. The Hoxd4b1hII construct is the Hoxd4 N-terminus including the first helix (1-180) fused commencing at the second helix of Hoxb1 (228-297) (Figure 3.6a). The converse Hoxb1d4hII construct is Hoxb1 (1-227) fused to Hoxd4 (181-250) (Figure 3.6a). *In vitro* pulldown analysis using these constructs revealed that Hoxd4b1hII, but not Hoxb1d4hII, interacted with Cdx1 (Figure 3.6b), strongly suggesting that sequences within helix I dictate association.

3.3.1 Identification of specific residues necessary for Hox-Cdx interaction

By comparing the sequences of the N-terminal arm and Helix I of Hox proteins capable of interacting with Cdx1 to the non-interactors revealed several residues potentially involved in this association (Figure 3.7). There are a number of residues in the first helix that are conserved, or semi-conserved, among all Hox proteins. Among the

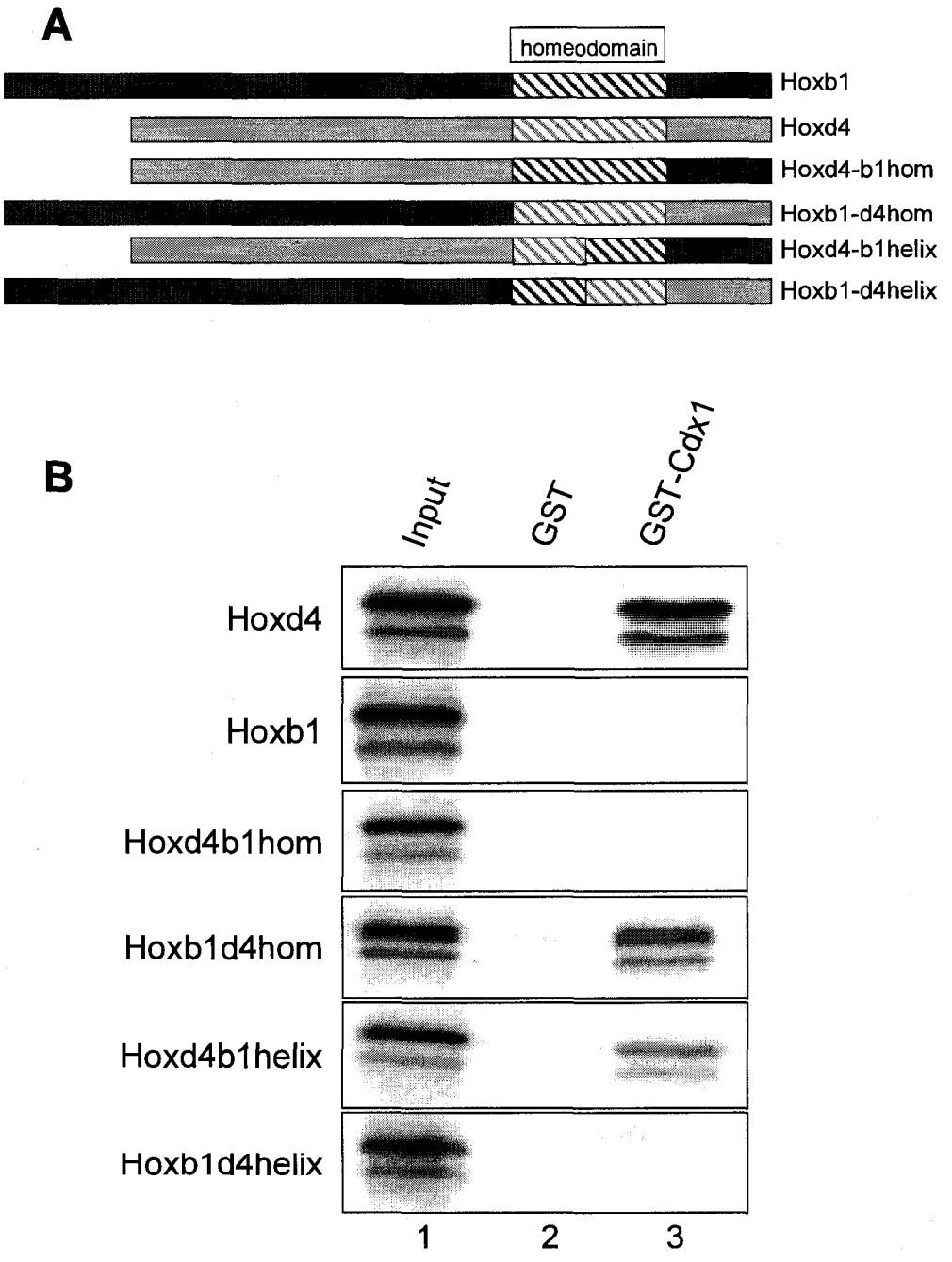

A
homeodomain
Hoxb1
Hoxd4
Hoxd4-b1hom
Hoxb1-d4hom
Hoxd4-b1helix
Hoxb1-d4helix
B
Input
GST
GST-Cdx1
Hoxd4
Hoxb1
Hoxd4b1hom
Hoxb1d4hom
Hoxd4b1helix
Hoxb1d4helix
1
2
3

Figure 3.6 – Interaction of Hoxb1-Hoxd4 fusion proteins with GST-Cdx1. The first helix of the homeodomain is sufficient for Hox binding to Cdx. A) Diagram representing the fusion constructs of Hoxb1 and Hoxd4 generated. B) Autoradiograph of ^{35}S labelled Hox proteins incubated with GST or GST-Cdx1. Hoxd4 and Hoxd4b1 interact with GST-Cdx1, while Hoxb1 and Hoxb1d4 do not. 20% ^{35}S labelled protein as input.

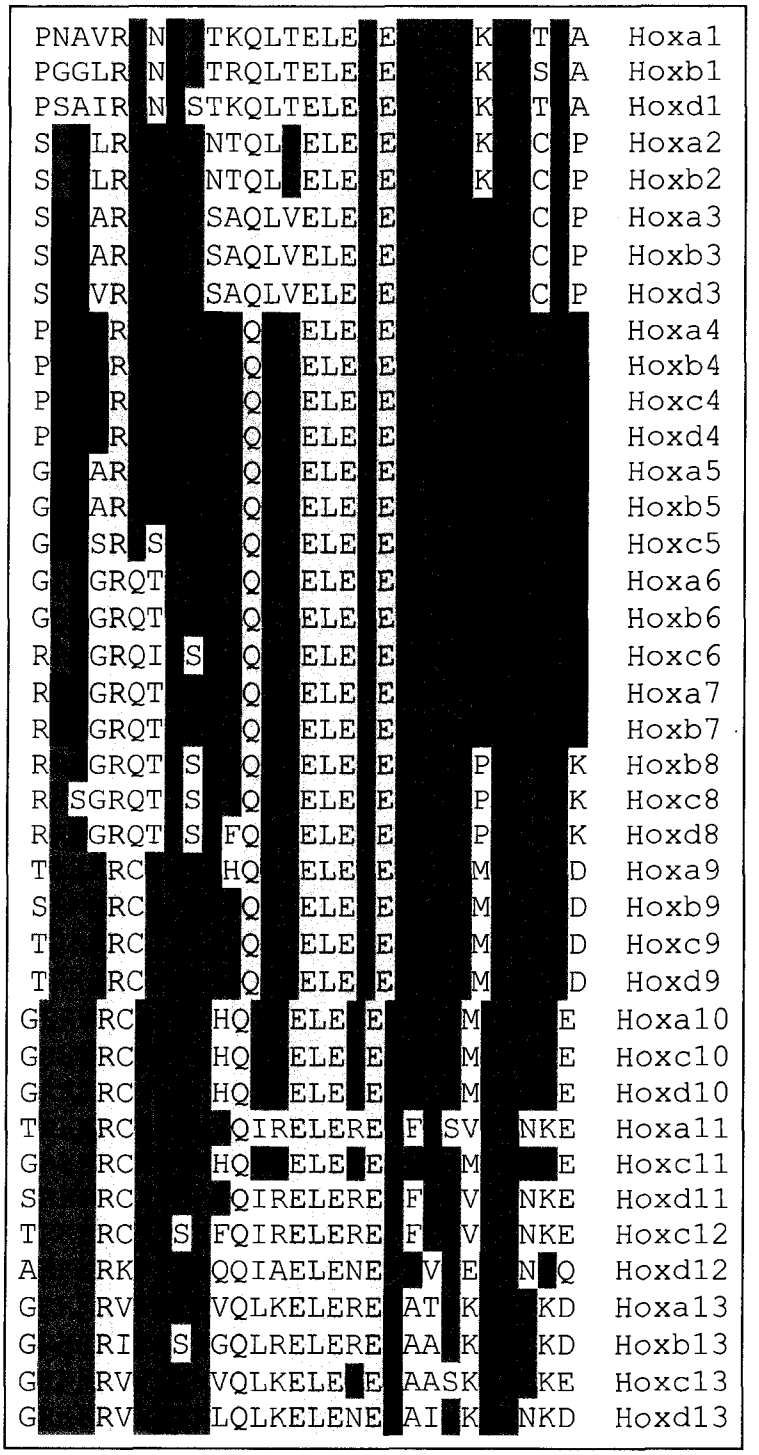

PNAVRNTKQLTELEEKTA Hoxa1
PGGLRNTRQLTELEEKSA Hoxb1
PSAIRNSTKQLTELEEKTA Hoxd1
SLRNTQLELEEKCP Hoxa2
SLRNTQLELEEKCP Hoxb2
SARSAQLVELEECP Hoxa3
SARSAQLVELEECP Hoxb3
SVRSAQLVELEECP Hoxd3
PRQELEE Hoxa4
PRQELEE Hoxb4
PRQELEE Hoxc4
PRQELEE Hoxd4
GARQELEE Hoxa5
GARQELEE Hoxb5
GSRSQELEE Hoxc5
GGRQTQELEE Hoxa6
GGRQTQELEE Hoxb6
RGRQISQELEE Hoxc6
RGRQTQELEE Hoxa7
RGRQTQELEE Hoxb7
RGRQTSQELEEPK Hoxb8
RSGRQTSQELEEPK Hoxc8
RGRQTSFQELEEPK Hoxd8
TRCHQELEEMD Hoxa9
SRCQELEEMD Hoxb9
TRCQELEEMD Hoxc9
TRCQELEEMD Hoxd9
GRCHQELEEME Hoxa10
GRCHQELEEME Hoxc10
GRCHQELEEME Hoxd10
TRCQIRELEREFSVNKE Hoxa11
GRCHQELEEME Hoxc11
SRCQIRELEREFVNKE Hoxd11
TRCSFQIRELEREFVNKE Hoxc12
ARKQQIAELENEVENQ Hoxd12
GRVVQLKELEREATKKD Hoxa13
GRISGQLRELEREAAKKD Hoxb13
GRVVQLKELEEAASKKE Hoxc13
GRVLQLKELENEAIKNKD Hoxd13

Figure 3.7 – Sequence alignment of Helix I. Yellow residues represent those conserved among all Hox proteins, turquoise residues among a majority of Hox, pink residues are conserved among more posterior proteins, and red residues among a smaller subset including group 4. Green residues represent those unique to group 4. Bright red residues are those used for mutagenesis of Hoxb1 to confer binding to Cdx1.

remaining residues, Hoxd4 contained arginines at positions 3, 24, and 29 of the homeodomain (residues 201, 222, 227 of full length Hoxb1 transcript). These arginine residues are also present in the other interactors, but in the non-interacting Hoxb1 and Hoxc8 homeodomains, these residues were neither arginine, nor were they conserved substitutions. In order to determine whether these residues are important for the Hoxd4-Cdx1 interaction, Hoxb1 was mutated so that one or more of these residues was changed to arginine. The single mutant $Hoxb1_{G\rightarrow R}$ changed the residue at position 3 from glycine to arginine, while the double mutant $Hoxb1_{K,A\rightarrow R,R}$ has mutations at positions 24 (lysine to arginine) and 29 (alanine to arginine) of the homeodomain. The triple mutant $Hoxb1_{G,K,A\rightarrow R,R,R}$ is a combination of the single and double mutants (Figure 3.8a).

In vitro pulldown analysis revealed that changing one residue, $Hoxb1_{G\rightarrow R}$, conferred interaction with GST-Cdx1. The double mutant $Hoxb1_{K,A\rightarrow R,R}$ also interacted, albeit relatively weakly, while the triple mutant $Hoxb1_{G,K,A\rightarrow R,R,R}$ also interacted with GST-Cdx1, as predicted from the single mutant (Figure 3.8b). These data indicate that these arginines, in particular at position 3, are important to confer binding to Cdx1. This result is supported by the fact that the Hoxd8 C-terminal is able to interact with Cdx1, while the Hoxc8 C-terminal is not. There are only two residue differences in the first helix: the arginine at position 3, and a phenylalanine to tyrosine at position 11 which is a semi-conserved substitution. Thus, arginines present in the first helix region of the homeodomain regulate interactions between Hox proteins and Cdx1/2.

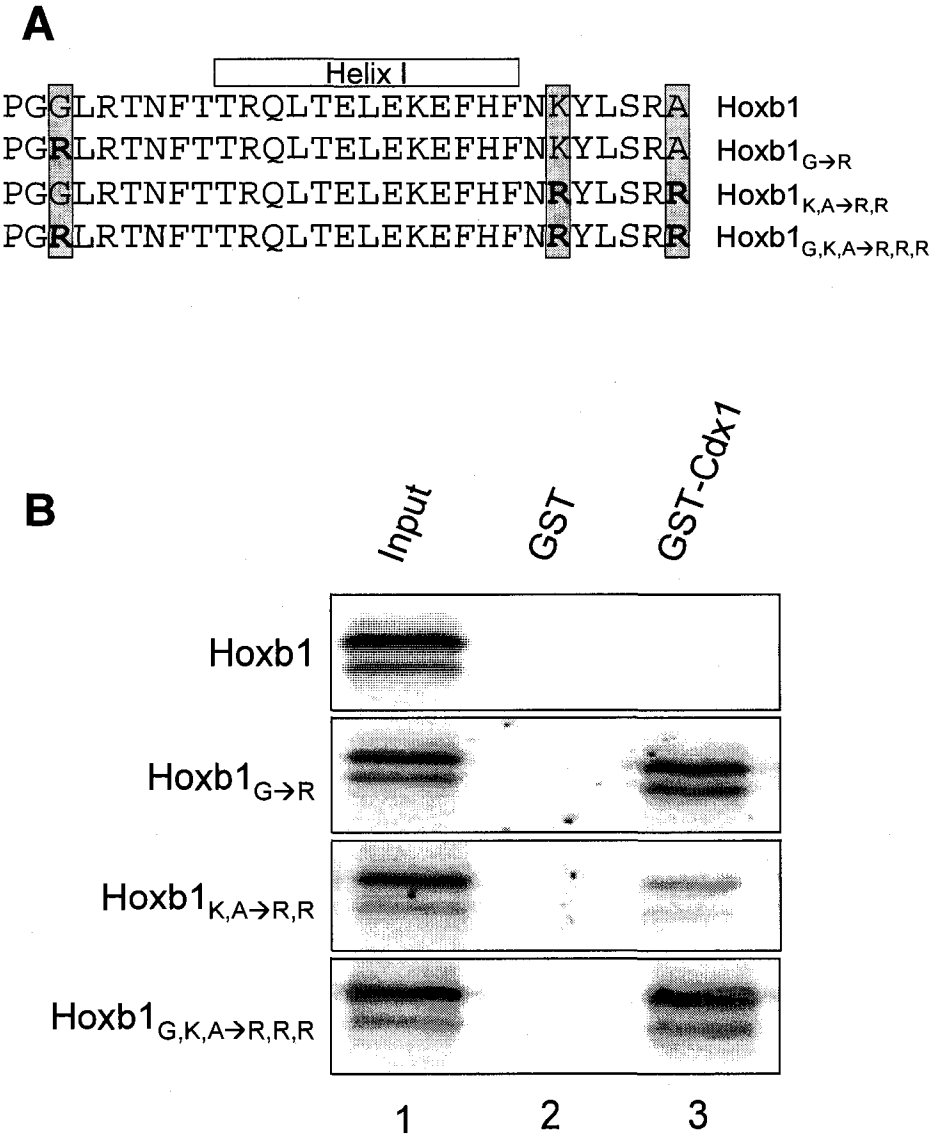

A
Helix I
PGGLRTNFTTRQLTELEKEFHFNKYLSRA Hoxb1
PGRLRTNFTTRQLTELEKEFHFNKYLSRA Hoxb1G→R
PGGLRTNFTTRQLTELEKEFHFNRYLSRR Hoxb1K,A→R,R
PGRLRTNFTTRQLTELEKEFHFNRYLSRR Hoxb1G,K,A→R,R,R
B
Input
GST
GST-Cdx1
Hoxb1
Hoxb1G→R
Hoxb1K,A→R,R
Hoxb1G,K,A→R,R,R
1
2
3

Figure 3.8 – Identification of residues sufficient to confer binding to Hoxb1. Autoradiograph of GST-Cdx1 pulldown of ^{35}S labelled Hox proteins. Hoxb1 was mutated to contain arginine residues present in Hoxd4. The triple mutant $Hoxb1_{G,K,A \rightarrow R,R,R}$ was able to bind with Cdx1, wheras the wildtype Hoxb1 was not. Input represents 20% of the ^{35}S labelled Hox proteins used.

3.4 The effects of a cellular environment on the interaction of Hox and Cdx

In vitro assays are useful to reveal direct protein-protein interactions, but do not necessarily reflect the cellular environment where these interactions presumably occur. Co-transfection assays, followed by co-immunoprecipitation (CO-IP) are the ideal experiments to assay whether proteins are interacting in a cellular environment. Unfortunately, although optimization of an immunoprecipitation protocol was attempted, a working CO-IP of Cdx and Hox was not attained. Thus, in order to assess the effects of a cellular environment on these interactions, GST-pulldowns were carried out with *Hox* expression vectors expressed in Cos7 cells. As specific Hox antibodies are not available, a FLAG epitope tag was integrated into the *Hox* constructs. The harvested cell lysate was then incubated with GST-Cdx and associated proteins assessed by Western blotting with anti-Flag. This analysis showed full-length Hoxb1, Hoxd4, Hoxa6 and Hoxa9 interact with both Cdx1 and Cdx2, while Hoxc8 did not interact with either protein (Figure 3.9), whereas in vitro, Hoxb1, Hoxa6, and Hoxa9 did not interact (Figure 3.4). Taken together, the translation of Hox proteins in a cellular environment affects the ability to bind with GST-Cdx proteins, due to the presence of necessary bridging proteins, or post-translational modifications.

3.4.1 Interaction of Cdx with other Hox partners

Certain Hox proteins that do not interact with GST-Cdx *in vitro*, or whose interactions are inhibited by N-terminal sequences, may associate with Cdx through common bridging factors. Hox proteins often associate with cofactors involved in DNA binding or subcellular distribution. Homeodomain proteins Meis1a and Pbx1 are two

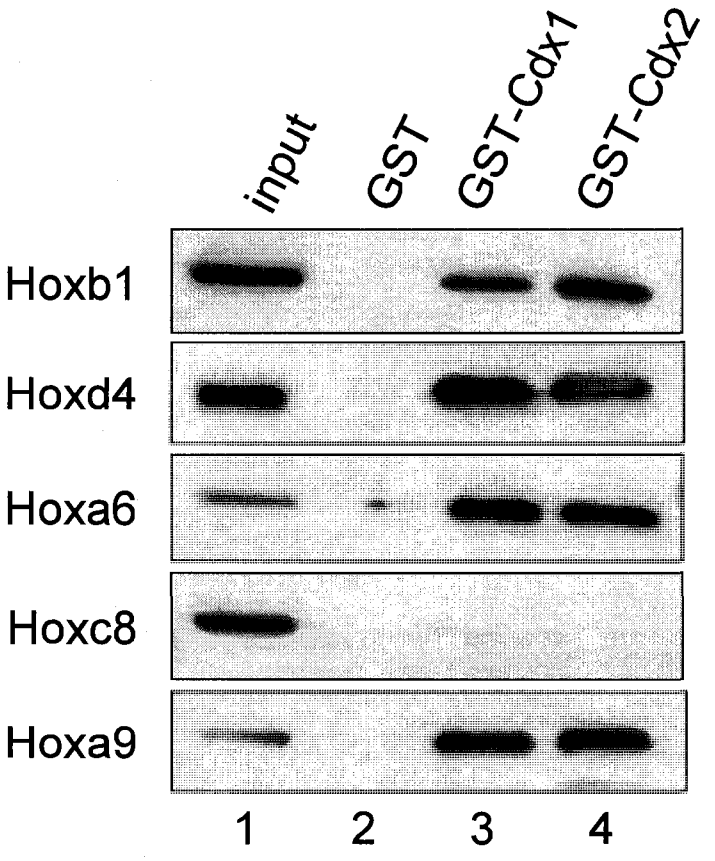

input
GST
GST-Cdx1
GST-Cdx2
Hoxb1
Hoxd4
Hoxa6
Hoxc8
Hoxa9
1
2
3
4

Figure 3.9 – Interaction of Hox-transfected cell lysates with GST-Cdx. Western blot of GST-Cdx pulldowns with protein extracts from transfected Cos7 cell lysates. Hoxb1, Hoxd4, Hoxa6, and Hoxa9 interact with both GST-Cdx1 and GST-Cdx2 while Hoxc8 does not. Input represents 5% of the lysate used.

such Hox-interacting proteins (Moens and Selleri 2006). As both are present in Cos7 cells, they appear to be excellent candidates for bridging proteins, and may explain the interactions between Cdx and Hox genes seen from lysates, but absent in *in vitro* assays. Indeed, consistent with this possibility, pulldown assays revealed that both Pbx1a and Meis1a can associate with GST-Cdx1 or Cdx2 (Figure 3.10).

3.5 Identification of possible targets for Hox-Cdx interaction

Given that a subset of Hox proteins interacts with Cdx1, the outstanding question is the biological significance of this interaction. One well documented role for both Cdx1 and Hox proteins is transcriptional regulation of *Hox* genes themselves (Popperl 1995; Subramanian et al. 1995; Charite et al. 1998; Studer et al. 1998). A number of *Hox* regulatory elements have already been characterized that contain both Hox and Cdx binding sites, and are of possible functional relevance to the Cdx-Hox complex. In this regard, an essential 200 bp regulatory element located upstream of *Hoxc8* (Shashikant et al. 1995) harbours Cdx and Hox binding sites that are two base pairs apart (Wang et al. 2004). To assess this element as a potential Cdx-Hox bipartite target, the enhancer sequence was used in transient transfection assays in P19 embryocarcinoma cells. As the Hox protein(s) that activates this element is unknown, Hoxc8 and Hoxd4 were both tested together with Cdx1. There was no synergistic activation, suggesting that this enhancer element is not affected by Cdx-Hox protein interactions.

The Hoxd4 autoregulatory element (ARE) was also surveyed for Cdx-Hox regulation, as there is a putative Cdx response element therein (Popperl and Featherstone 1993), and the Hox protein needed to drive transcription (Hoxd4) of this element is

known. Inconsistent results with this element rendered potential Hox-Cdx1 interactions on this motif difficult to assess (data not shown).

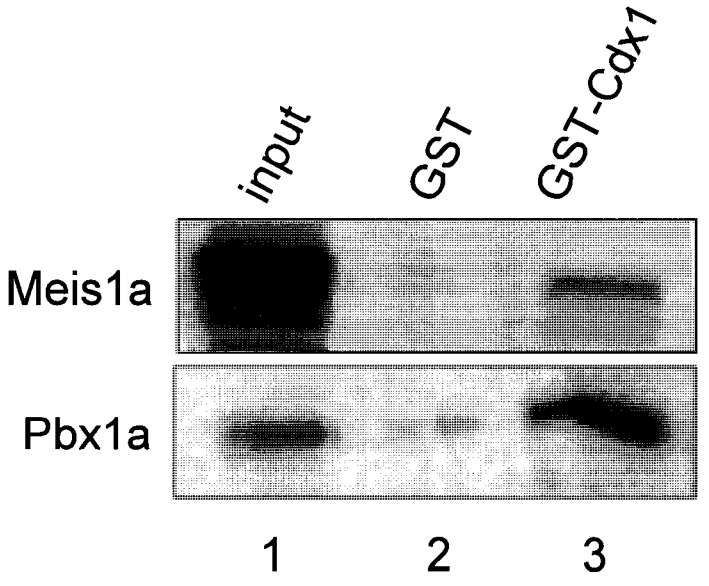

input
GST
GST-Cdx1
Meis1a
Pbx1a
1
2
3

Figure 3.10 – Interaction of Pbx and Meis with GST-Cdx. Autoradiograph of ^{35}S labelled homeodomain proteins Meis1a and Pbx1a pulldown with GST or GST-Cdx1. Both proteins interact with Cdx1, while Meis1a also interacts with Cdx2. Input represents 20% of the ^{35}S labelled Hox proteins used.

Chapter 4

Discussion

Physical interactions have been described for a number of transcription factors, and are well known to impact their function through, for example, influencing binding specificity or avidity or subcellular localization with subsequent effects on target gene expression (Moens and Selleri 2006). Among other roles, both Cdx and Hox members are involved in patterning the AP axis, and both have previously documented binding partners. For example, Cdx proteins have been reported to interact with other homeodomain proteins such as Pbx1, Brn-4, and HNF1alpha (Boudreau et al. 2002; Liu et al. 2006; Wang et al. 2006), while Hox proteins have been shown to interact with homeodomain protein partners such as members of the Pbx and Meis families (Moens and Selleri 2006).

Cdx members are direct regulators of expression of a number of *Hox* genes which in turn regulate cohorts of largely unknown downstream genes involved in AP patterning. As many Hox proteins require interaction with homeodomain-containing cofactors such as Pbx and Meis family members to execute their transcriptional function, Cdx proteins were hypothesized to also act as Hox cofactors, in addition to driving *Hox* expression. Indeed, Hox proteins are known to auto- and cross-regulate, and thus a Hox-Cdx interaction may be involved in governing *Hox* expression.

In this thesis I present data showing that Cdx1 and Cdx2 interact directly with a subset of Hox proteins. This interaction is dependent on specific residues in the first helix of the Hox homeodomain, and influenced by unknown sequences in the N-terminus. This binding profile is, however, altered in Hox proteins isolated from cells, suggesting the involvement of additional binding factors. To this end, I show that Cdx is also capable of interacting with the Hox cofactors Pbx1a and Meis1a, implying that Cdx could

be involved in a complex that includes Hox and other binding partners. The *in vivo* relevance of these interactions was investigated by testing the enhancer regions of various *Hox* genes that harbour both Cdx and Hox binding sites, but no conclusive regulatory interactions were observed. This latter observation notwithstanding, these studies extend our understanding of transcription factor interactions, and are the first description of a Hox-Cdx complex.

4.1 Colocalization of Hox and Cdx

Although transcription factors bind DNA, some are sequestered in the cytoplasm prior to receiving signals that cause their translocation to the nucleus (Saleh et al. 2000). In this regard, both Cdx1 and Hoxd4 colocalize to the nucleus in P19 and Cos7 cell lines (Figure 3.2), indicating that an interaction can physically occur between these transcription factors in a cellular environment. *Hox* and *Cdx* genes also exhibit similar expression patterns *in vivo*, being expressed in the primitive streak and in all three germ layers in the caudal embryo, and both classes exert their effects on vertebral patterning in this expression domain. Although Cdx proteins function upstream of *Hox* family members, this does not preclude an additional role as a co-factor, perhaps in a manner similar to Cdx-LEF1 interactions on the *Cdx1* promoter (Beland et al. 2004).

4.2 A subset of Hox proteins interact with Cdx1

Preliminary experiments indicated that Cdx1 interacts with Hoxd4, but not Hoxa9, and that the interaction is localized to the C-terminus of both proteins. This

region includes the highly conserved homeodomain, but not the Hox hexapeptide motif involved in binding Pbx and Meis (Chang et al. 1995).

Although Hox proteins are highly conserved at the level of the homeodomain, there are differences in the motif between paralog groups and to a lesser extent within a given paralog group. To this effect, Hox C-terminus fragments were used in order to further define the region of interaction, employing Hox proteins representative of the different paralog groups to determine the extent of the Cdx-Hox interaction. These studies defined the general domain of interaction as the homeodomain for both proteins, while analysis of the panel of Hox proteins examined led to identification of interactors and non-interactors, which could be used to predict specific residues required for interaction. In this regard, there is ample prior evidence implicating the homeodomain from diverse proteins as an interface in association with other proteins (Bendall et al. 1998; Kawata et al. 2003; Zaffran and Frasch 2005). As the homeodomain is also the site of DNA binding, such interactions could potentially enhance or exclude DNA binding, depending on the nature of the interface.

Interestingly, while most Hox C-termini tested interact with Cdx1 *in vitro*, certain of the corresponding full length proteins did not (Figures 3.3, 3.4). For example, Hoxa6 did not interact as a full length protein, but did as a homeodomain fragment. Through a series of domain swapping experiments, the N-terminus of Hoxa6 was shown to inhibit the binding to Cdx1, and this inhibitory function could be conveyed to other interacting homeodomains. For example, fusion of the Hoxa6 N-terminus with the C-terminus of Hoxd4 results in a lack of binding with GST-Cdx1. However, the opposite fusion, Hoxd4 N-terminus with Hoxa6 C-terminus, is still capable of binding GST-Cdx1 (Figure

3.5). This inhibitory function may reflect a level of specificity in which protein interactions involving the highly conserved homeodomain are controlled, further modulating Hox-Cdx interactions. Additional considerations may also come into play in the context of the cellular environment, as discussed in further detail below.

4.3 Characterization of Hox homeodomain residues necessary for Cdx interaction

While the N-terminus of some Hox proteins can modulate interaction with Cdx, the homeodomain itself must be able to bind Cdx alone for the interaction to occur as either a C-terminal fragment or in the context of the full length protein. To identify the region of Hox proteins, and specific residues, involved in this interaction, comparisons were made between the homeodomains of an interactor (Hoxd4) versus a non-interactor (Hoxb1). In addition, chimeric proteins were generated in which regions of the homeodomains were exchanged between interactors and non-interactors, leading to the conclusion that the N-terminal arm and helix I region of the homeodomain is involved in binding to Cdx1. Closer inspection of this region yielded a number of interesting residues; there were several arginines present in the homeodomain of Hoxd4 and other positive interactors while the same residues harboured non-conserved substitutions in the non-interactors. Protein-protein interactions are often mediated by polar residues. Arginine, a basic amino acid, can be involved in binding negatively charged groups and is one of several amino acidy typically enriched in protein-protein interfaces (Bogan and Thorn 1998; Jones et al. 2000). Thus, arginine is a prime candidate for being the site of Hox-Cdx1 interactions.

In assessing the potential role of the above arginines in mediating Hox-Cdx interactions, it was found that the mutation of a single glycine to arginine in the third position of the homeodomain was sufficient to confer binding onto Hoxb1, a non-interacting protein. In this regard, the homeodomains of most Hox paralog groups 1-8 have an arginine in position 3 while groups 9-13 have a lysine, a semi-conserved substitution; no paralog group 1 member has an arginine in this position, correlating with binding to Cdx1. In addition, Hoxc8 has a non-conserved substitution at the third residue of the homeodomain, while Hoxd8 has an arginine at this position, in following with the data that Hoxd8 interacted with GST-Cdx1, while Hoxc8 did not. Interestingly, Hoxc8 does not express as far rostral as Hoxd8 and Hoxb8 do in the neuroectoderm/neural tube, and this could be indicative of differential regulation or functional differences within a paralog group. Therefore, the third residue of the homeodomain of Hox proteins seems to be important in conveying ability to bind Cdx1.

Previous studies have shown that the N-terminal arm of the homeodomain is important for interactions both with DNA as well as accessory proteins (Kissinger et al. 1990; Phelan and Featherstone 1997). There are two regions of the homeodomain that make physical contact with DNA: the third helix and the N-terminal arm (Otting et al. 1990; Wolberger et al. 1991). Residues 2 and 3 of the homeodomain are responsible for distinguishing binding of Hoxa1 and Hoxd4 to DNA, as monomers (Phelan et al. 1994). The interaction of the N-terminal arm of the Hox homeodomain with DNA is also affected by the Hox-Pbx interaction, and may be important in conferring functional specificity among Hox proteins (Phelan and Featherstone 1997). Taken together, this

data shows that this region of Hox proteins is important for protein-protein interactions, in following with the interaction between Hox and Cdx.

4.4 Cellular Environment and Hox Cofactors

As opposed to GST pulldowns using *in vitro* synthesized Hox proteins, all Hox proteins, with the exception of Hoxc8, generated in a cellular environment interacted with GST-Cdx1 *in vitro*. This finding may be due to one or more of a number of reasons; as these proteins are transcription factors, the Hox-Cdx interaction might only have an effect when bound to target DNA sequences; a bridging protein present in Cos7 cells could bind both Hox and Cdx, so they do not directly contact each other but are part of a complex; finally post-translational modification (eg. phosphorylation, protein folding) of Hox proteins could affect their interaction with Cdx1.

The well documented Hox-binding homeodomain proteins Meis1a and Pbx1 are present in Cos7 cells (Moens and Selleri, 2006). Moreover, I have shown that both of these proteins could bind to Cdx1 or Cdx2 *in vitro* (Figure 3.10). This finding suggests that Pbx and/or Meis could be involved as bridging factors which are required for some of the Cdx-Hox interactions seen using Cos7 extracts. The inhibitory N-terminal regions of Hoxa6 that are observed *in vitro*, but not when expressed in Cos7 cells, may also be explained by interaction of Hoxa6 and Cdx1 with cofactors. The addition of a third protein may induce a conformational change. This is reminiscent of the change from inhibition to activation when Hox and Pbx bound to DNA are exposed to different levels of co-activators and co-repressors such as CBP and HDAC (Saleh et al. 2000).

4.5 Functional significance of Cdx-Hox complexes

DNA binding sites for Hox proteins are difficult to predict due to the highly divergent sequences to which they can bind (Ebner et al. 2005), and the number of genes directly regulated by Hox are thought to be more numerous than initially assumed (Williams et al. 2005b). For this reason, targets of the Hox-Cdx interaction were first investigated within the Hox loci, as *Hox* genes are known to auto- and cross-regulate, while Cdx is upstream of a number of *Hox* genes (Popperl 1995; Subramanian et al. 1995; Charite et al. 1998; Studer et al. 1998).

The *Hoxc8* early enhancer region harbours both Cdx and Hox binding sites in close proximity, making is a logical transcriptional target. However, a reporter construct harboring these elements did not exhibit synergistic interactions in response to Cdx1 and Hoxc8 or Hoxd4. In this regard, Hoxc8 does not bind Cdx1 under any circumstance tested, and the resulting lack of synergy was expected. Hoxd4 and Cdx1 do interact under all circumstances, but did not activate this enhancer region. It is, however, possible that Hoxd4 is not the Hox protein required to regulate *Hoxc8*.

The *Hoxd4* autoregulatory element (Hoxd4ARE) has both Cdx and Hox binding sites, and Hoxd4 is known to transcriptionally regulate this element (Popperl and Featherstone 1992). Although both Cdx1 and Hoxd4 activated this enhancer in transfection assays, no conclusive synergy was exhibited between these transcription factors (data not shown). This could be due to the possibility that the Cdx-Hox interaction requires other cofactors, such as Pbx and Meis, to be fully functional.

It is possible that Cdx binds and affects Hox function without binding to the DNA itself, as seen previously with the Cdx-LEF interaction on the Cdx1 promoter (Beland et

al. 2004). This would suggest the presence of both Cdx and Hox binding sites in a putative target gene promoter is unnecessary, although the Hox binding sites tested did not show activation of transcription, arguing against this theory. There is no true way to find out how Cdx and Hox interact at the protein level to affect target gene expression until relevant genes and their promoter regions have been identified.

4.6 Future Directions

Co-transfections in Cos7 cells, followed by co-immunoprecipitation would be ideal to show that this interaction is capable of occurring entirely in a cellular environment, and thus potentially in the developing embryo. Co-immunoprecipitations could also elucidate the role that Pbx and Meis have in the Cdx-Hox interaction, whether they are present in all cellular Cdx-Hox interactions or only when needed as bridging factors.

The biological significance of these interactions remains to be confirmed. In order to identify target sequences bound by the Hox-Cdx protein complex, chromatin immunoprecipitation (ChIP) on a microarray chip, or ChIP-on-chip, can be performed. The source material required would need to be tissue that co-expresses Hox and Cdx, such as the paraxial mesoderm of the developing embryo. A double immunoprecipitation, with both Cdx and Hox specific antibodies, would need to be employed such that the target sequences retrieved by this experiment are associated with both Cdx and Hox proteins. Assaying the purified DNA on a microarray would potentially show where these complexes bind DNA, and thus target genes. Another alternative is ChIP-PET, or chromatin immunoprecipitation followed by pair end ditag sequencing (Wei et al. 2006). DNA fragments acquired by ChIP (in the same manner as

the ChIP-on-chip) are tagged by a known sequence on both ends. These ditagged fragments are then linked together and sequenced, again resulting in identification of potential target genes. These approaches have an advantage over bioinformatical approaches, as they do not require knowledge of how the Cdx-Hox interaction binds DNA, but are highly technically demanding, due to the necessity of a double immunoprecipitation.

Once target sequences and their associated genes are identified, the binding sites of the Hox-Cdx interaction can be described using luciferase assays and deletion mutants. Confirmation of *in vitro* binding of these proteins to the DNA fragments can be carried out using an electrophoretic mobility shift assay (EMSA), while *in vivo* relevance could be confirmed by ChIP. Functional activity of these putative regions, by Cdx and Hox in conjunction, can then be investigated by luciferase assays and possibly transgenic mice.

4.7 Conclusions

Hox-Cdx protein interactions appear to be multi-faceted, with Cdx1 able to bind directly to some, and requiring cell specific settings for other, Hox proteins. An arginine as the third residue of the Hox homeodomain is sufficient for binding to Cdx1, and this interaction can be mediated by inhibitory sequences in the N-terminal region of the Hox protein. The biological relevance of this interaction has yet to be deciphered. As they are both transcription factors, and function to pattern the AP axis during development, I would speculate that the interaction could be important in modulating downstream gene expression. This work has, and will continue to advance our understanding of the transcription network involved in patterning the AP axis.

Chapter 5

References

Barrow, J.R. 2006. Wnt/PCP signaling: a veritable polar star in establishing patterns of polarity in embryonic tissues. *Semin Cell Dev Biol* **17**(2): 185-193.

Barrow, J.R., Thomas, K.R., Boussadia-Zahui, O., Moore, R., and Kemler, R. 2003. Ectodermal Wnt3/beta-catenin signaling is required for the establishment and maintenance of the apical ectodermal ridge. *Genes Dev* **17**: 394.

Beck, F., Chawengsaksophak, K., Waring, P., Playford, R.J., and Furness, J.B. 1999. Reprogramming of intestinal differentiation and intercalary regeneration in Cdx2 mutant mice. *Proc Natl Acad Sci U S A* **96**(13): 7318-7323.

Beck, F., Erler, T., Russell, A., and James, R. 1995. Expression of Cdx-2 in the mouse embryo and placenta: Possible role in patterning of the extra-embryonic membranes. *Developmental Dynamics* **204**(3): 219-227.

Bel-Vialar, S., Itasaki, N., and Krumlauf, R. 2002. Initiating Hox gene expression: in the early chick neural tube differential sensitivity to FGF and RA signaling subdivides the HoxB genes in two distinct groups. *Development* **129**(22): 5103-5115.

Beland, M., Pilon, N., Houle, M., Oh, K., Sylvestre, J.R., Prinos, P., and Lohnes, D. 2004. Cdx1 autoregulation is governed by a novel Cdx1-LEF1 transcription complex. *Mol Cell Biol* **24**(11): 5028-5038.

Bendall, A.J., Rincon-Limas, D.E., Botas, J., and Abate-Shen, C. 1998. Protein complex formation between Msx1 and Lhx2 homeoproteins is incompatible with DNA binding activity. *Differentiation* **63**(3): 151-157.

Berthelsen, J., Kilstrup-Nielsen, C., Blasi, F., Mavilio, F., and Zappavigna, V. 1999. The subcellular localization of PBX1 and EXD proteins depends on nuclear import

and export signals and is modulated by association with PREP1 and HTH. *Genes Dev* **13**(8): 946-953.

Berthelsen, J., Zappavigna, V., Mavilio, F., and Blasi, F. 1998. Prep1, a novel functional partner of Pbx proteins. *EMBO J* **17**(5): 1423-1433.

Bogan, A.A. and Thorn, K.S. 1998. Anatomy of hot spots in protein interfaces. *J Mol Biol* **280**(1): 1-9.

Bottcher, R.T. and Niehrs, C. 2005. Fibroblast Growth Factor Signaling during Early Vertebrate Development. *Endocr Rev* **26**(1): 63-77.

Boudreau, F., Rings, E.H., van Wering, H.M., Kim, R.K., Swain, G.P., Krasinski, S.D., Moffett, J., Grand, R.J., Suh, E.R., and Traber, P.G. 2002. Hepatocyte nuclear factor-1 alpha, GATA-4, and caudal related homeodomain protein Cdx2 interact functionally to modulate intestinal gene transcription. Implication for the developmental regulation of the sucrase-isomaltase gene. *J Biol Chem* **277**(35): 31909-31917.

Brooke, N.M., Garcia-Fernandez, J., and Holland, P.W. 1998. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Nature* **392**(6679): 920-922.

Calon, A., Gross, I., Davidson, I., Kedinger, M., Duluc, I., Domon-Dell, C., and Freund, J.N. 2007. Functional interaction between the homeoprotein CDX1 and the transcriptional machinery containing the TATA-binding protein. *Nucleic Acids Res* **35**(1): 175-185.

Chang, C.P. 1995. Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. *Genes Dev* **9**: 663-674.

Chang, C.P., Jacobs, Y., Nakamura, T., Jenkins, N.A., Copeland, N.G., and Cleary, M.L. 1997. Meis proteins are major in vivo DNA binding partners for wild-type but not chimeric Pbx proteins. *Mol Cell Biol* **17**(10): 5679-5687.

Chang, C.P., Shen, W.F., Rozenfeld, S., Lawrence, H.J., Largman, C., and Cleary, M.L. 1995. Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. *Genes Dev* **9**(6): 663-674.

Charite, J., de Graaff, W., Consten, D., Reijnen, M.J., Korving, J., and Deschamps, J. 1998. Transducing positional information to the Hox genes: critical interaction of cdx gene products with position-sensitive regulatory elements. *Development* **125**(22): 4349-4358.

Ciruna, B. and Rossant, J. 2001. FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. *Dev Cell* **1**: 37-49.

Condie, B.G. and Capecchi, M.R. 1994. Mice with targeted disruptions in the paralogous genes hoxa-3 and hoxd-3 reveal synergistic interactions. *Nature* **370**(6487): 304-307.

Cook, C.E., Jimenez, E., Akam, M., and Salo, E. 2004. The Hox gene complement of acoel flatworms, a basal bilaterian clade. *Evol Dev* **6**(3): 154-163.

Crossley, P.H. and Martin, G.R. 1995. The mouse Fgf8 gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo. *Development* **121**(2): 439-451.

Dubrulle, J. and Pourquie, O. 2004. Coupling segmentation to axis formation. *Development* **131**(23): 5783-5793.

Duprey, P., Chowdhury, K., Dressler, G.R., Balling, R., Simon, D., Guenet, J.L., and Gruss, P. 1988. A mouse gene homologous to the Drosophila gene caudal is expressed in epithelial cells from the embryonic intestine. *Genes Dev* **2**(12A): 1647-1654.

Ebner, A., Cabernard, C., Affolter, M., and Merabet, S. 2005. Recognition of distinct target sites by a unique Labial/Extradenticle/Homothorax complex. *Development* **132**(7): 1591-1600.

Epstein, M., Pillemer, G., Yelin, R., Yisraeli, J.K., and Fainsod, A. 1997. Patterning of the embryo along the anterior-posterior axis: the role of the caudal genes. *Development* **124**(19): 3805-3814.

Galant, R., Walsh, C.M., and Carroll, S.B. 2002. Hox repression of a target gene: extradenticle-independent, additive action through multiple monomer binding sites. *Development* **129**: 3115-3126.

Gaunt, S.J., Cockley, A., and Drage, D. 2004. Additional enhancer copies, with intact cdx binding sites, anteriorize Hoxa-7/lacZ expression in mouse embryos: evidence in keeping with an instructional cdx gradient. *Int J Dev Biol* **48**(7): 613-622.

Gautier-Stein, A., Domon-Dell, C., Calon, A., Bady, I., Freund, J.N., Mithieux, G., and Rajas, F. 2003. Differential regulation of the glucose-6-phosphatase TATA box by intestine-specific homeodomain proteins CDX1 and CDX2. *Nucleic Acids Res* **31**(18): 5238-5246.

Gehring, W.J., Qian, Y.Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A.F., Resendez-Perez, D., Affolter, M., Otting, G., and Wuthrich, K. 1994. Homeodomain-DNA recognition. *Cell* **78**(2): 211-223.

Glover, J.C., Renaud, J.S., and Rijli, F.M. 2006. Retinoic acid and hindbrain patterning. *J Neurobiol* **66**(7): 705-725.

Goutte, C. and Johnson, A.D. 1994. Recognition of a DNA operator by a dimer composed of two different homeodomain proteins. *EMBO J* **13**(6): 1434-1442.

Green, S., Issemann, I., and Sheer, E. 1988. A versatile in vivo and in vitro eukaryotic expression vector for protein engineering. *Nucleic Acids Res* **16**(1): 369.

Haremaki, T., Tanaka, Y., Hongo, I., Yuge, M., and Okamoto, H. 2003. Integration of multiple signal transducing pathways on Fgf response elements of the Xenopus caudal homologue Xcad3. *Development* **130**(20): 4907-4917.

Hoey, T. and Levine, M. 1988. Divergent homeo box proteins recognize similar DNA sequences in Drosophila. *Nature* **332**(6167): 858-861.

Horn, J.M. and Ashworth, A. 1995. A member of the caudal family of homeobox genes maps to the X-inactivation centre region of the mouse and human X chromosomes. *Hum Mol Genet* **4**(6): 1041-1047.

Houle, M., Sylvestre, J.R., and Lohnes, D. 2003. Retinoic acid regulates a subset of Cdx1 function in vivo. *Development* **130**(26): 6555-6567.

Hussain, M.A. and Habener, J.F. 1999. Glucagon gene transcription activation mediated by synergistic interactions of pax-6 and cdx-2 with the p300 co-activator. *J Biol Chem* **274**(41): 28950-28957.

Iimura, T. and Pourquie, O. 2006. Collinear activation of Hoxb genes during gastrulation is linked to mesoderm cell ingression. *Nature* **442**(7102): 568-571.

-. 2007. Hox genes in time and space during vertebrate body formation. *Development, Growth & Differentiation* **49**(4): 265-275.

Isaacs, H.V., Pownall, M.E., and Slack, J.M. 1998. Regulation of Hox gene expression and posterior development by the Xenopus caudal homologue Xcad3. *EMBO J* **17**(12): 3413-3427.

Jacobs, Y., Schnabel, C.A., and Cleary, M.L. 1999. Trimeric association of Hox and TALE homeodomain proteins mediates Hoxb2 hindbrain enhancer activity. *Mol Cell Biol* **19**(7): 5134-5142.

James, R., Erler, T., and Kazenwadel, J. 1994. Structure of the murine homeobox gene cdx-2. Expression in embryonic and adult intestinal epithelium. *J Biol Chem* **269**(21): 15229-15237.

Jaw, T.J., You, L.R., Knoepfler, P.S., Yao, L.C., Pai, C.Y., Tang, C.Y., Chang, L.P., Berthelsen, J., Blasi, F., Kamps, M.P., and Sun, Y.H. 2000. Direct interaction of two homeoproteins, homothorax and extradenticle, is essential for EXD nuclear localization and function. *Mech Dev* **91**(1-2): 279-291.

Jones, S., Marin, A., and Thornton, J.M. 2000. Protein domain interfaces: characterization and comparison with oligomeric protein interfaces. *Protein Eng* **13**(2): 77-82.

Kawata, H., Yamada, K., Shou, Z., Mizutani, T., and Miyamoto, K. 2003. The mouse zinc-fingers and homeoboxes (ZHX) family; ZHX2 forms a heterodimer with ZHX3. *Gene* **323**: 133-140.

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**(3): 579-590.

Kmita, M. and Duboule, D. 2003. Organizing axes in time and space; 25 years of colinear tinkering. *Science* **301**(5631): 331-333.

Knittel, T., Kessel, M., Kim, M.H., and Gruss, P. 1995. A conserved enhancer of the human and murine Hoxa-7 gene specifies the anterior boundary of expression during embryonal development. *Development* **121**(4): 1077-1088.

Kostic, D. and Capecchi, M.R. 1994. Targeted disruptions of the murine Hoxa-4 and Hoxa-6 genes result in homeotic transformations of components of the vertebral column. *Mech Dev* **46**(3): 231-247.

Lickert, H. and Kemler, R. 2002. Functional analysis of cis-regulatory elements controlling initiation and maintenance of early Cdx1 gene expression in the mouse. *Dev Dyn* **225**(2): 216-220.

Liu, P., Wakamiya, M., Shea, M.J., Albrecht, U., Behringer, R.R., and Bradley, A. 1999. Requirement for Wnt3 in vertebrate axis formation. *Nat Genet* **22**: 361.

Liu, T., Branch, D.R., and Jin, T. 2006. Pbx1 is a co-factor for Cdx-2 in regulating proglucagon gene expression in pancreatic A cells. *Mol Cell Endocrinol* **249**(1-2): 140-149.

Logan, C.Y. and Nusse, R. 2004. THE WNT SIGNALING PATHWAY IN DEVELOPMENT AND DISEASE. *Annual Review of Cell and Developmental Biology* **20**(1): 781-810.

Lohnes, D. 2003. The Cdx1 homeodomain protein: an integrator of posterior signaling in the mouse. *Bioessays* **25**(10): 971-980.

Lohnes, D., Mark, M., Mendelsohn, C., Dolle, P., Decimo, D., LeMeur, M., Dierich, A., Gorry, P., and Chambon, P. 1995. Developmental roles of the retinoic acid receptors. *J Steroid Biochem Mol Biol* **53**(1-6): 475-486.

Longobardi, E. and Blasi, F. 2003. Overexpression of PREP-1 in F9 teratocarcinoma cells leads to a functionally relevant increase of PBX-2 by preventing its degradation. *J Biol Chem* **278**(40): 39235-39241.

Mann, R.S. and Affolter, M. 1998. Hox proteins meet more partners. *Curr Opin Genet Dev* **8**(4): 423-429.

Margalit, Y., Yarus, S., Shapira, E., Gruenbaum, Y., and Fainsod, A. 1993. Isolation and characterization of target sequences of the chicken CdxA homeobox gene. *Nucleic Acids Res* **21**(21): 4915-4922.

Mead, P.E., Brivanlou, I.H., Kelley, C.M., and Zon, L.I. 1996. BMP-4-responsive regulation of dorsal-ventral patterning by the homeobox protein Mix.1. *Nature* **382**(6589): 357-360.

Merabet, S., Kambris, Z., Capovilla, M., Berenger, H., Pradel, J., and Graba, Y. 2003. The hexapeptide and linker regions of the AbdA Hox protein regulate its activating and repressive functions. *Dev Cell* **4**(5): 761-768.

Meyer, B.I. and Gruss, P. 1993. Mouse Cdx-1 expression during gastrulation. *Development* **117**(1): 191-203.

Mitchelmore, C., Troelsen, J.T., Spodsberg, N., Sjostrom, H., and Noren, O. 2000. Interaction between the homeodomain proteins Cdx2 and HNF1alpha mediates expression of the lactase-phlorizin hydrolase gene. *Biochem J* **346 Pt 2**: 529-535.

Moens, C.B. and Selleri, L. 2006. Hox cofactors in vertebrate development. *Dev Biol* **291**(2): 193-206.

Moghadam, H.K., Ferguson, M.M., and Danzmann, R.G. 2005. Evidence for Hox gene duplication in rainbow trout (Oncorhynchus mykiss): a tetraploid model species. *J Mol Evol* **61**(6): 804-818.

Moreno, E. and Morata, G. 1999. Caudal is the Hox gene that specifies the most posterior Drosophile segment. *Nature* **400**(6747): 873-877.

Niederreither, K., McCaffery, P., Drager, U.C., Chambon, P., and Dolle, P. 1997. Restricted expression and retinoic acid-induced downregulation of the retinaldehyde dehydrogenase type 2 (RALDH-2) gene during mouse development. *Mech Dev* **62**(1): 67-78.

Nieuwkoop, P.D. 1952. Activation and Organization of the Central Nervous System in Amphibians .3. Synthesis of a New Working Hypothesis. *Journal of Experimental Zoology* **120**(1): 83-108.

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**(10): 3085-3092.

Pearson, J.C., Lemons, D., and McGinnis, W. 2005. Modulating Hox gene functions during animal body patterning. *Nat Rev Genet* **6**(12): 893-904.

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**(13): 8635-8643.

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**(8): 5066-5075.

Pilon, N., Oh, K., Sylvestre, J.R., Savory, J.G., and Lohnes, D. 2007. Wnt signaling is a key mediator of Cdx1 expression in vivo. *Development* **134**(12): 2315-2323.

Poellinger, L., Yoza, B.K., and Roeder, R.G. 1989. Functional cooperativity between protein molecules bound at two distinct sequence elements of the immunoglobulin heavy-chain promoter. *Nature* **337**(6207): 573-576.

Popperl, H. 1995. Segmental expression of Hoxb-1 is controlled by a highly conserved autoregulatory loop dependent upon Exd/Pbx. *Cell* **81**: 1031-1042.

Popperl, H. and Featherstone, M.S. 1992. An autoregulatory element of the murine Hox-4.2 gene. *EMBO J* **11**(10): 3673-3680.

-. 1993. Identification of a retinoic acid response element upstream of the murine Hox-4.2 gene. *Mol Cell Biol* **13**(1): 257-265.

Postlethwait, J.H. and Schneiderman, H.A. 1969. A clonal analysis of determination in Antennapedia a homoeotic mutant of Drosophila melanogaster. *Proc Natl Acad Sci U S A* **64**(1): 176-183.

Pownall, M.E., Tucker, A.S., Slack, J.M., and Isaacs, H.V. 1996. eFGF, Xcad3 and Hox genes form a molecular pathway that establishes the anteroposterior axis in Xenopus. *Development* **122**(12): 3881-3892.

Prinos, P., Joseph, S., Oh, K., Meyer, B.I., Gruss, P., and Lohnes, D. 2001. Multiple pathways governing Cdx1 expression during murine development. *Dev Biol* **239**(2): 257-269.

Remacle, S., Abbas, L., De Backer, O., Pacico, N., Gavalas, A., Gofflot, F., Picard, J.J., and Rezsohazy, R. 2004. Loss of function but no gain of function caused by amino acid substitutions in the hexapeptide of Hoxa1 in vivo. *Mol Cell Biol* **24**(19): 8567-8575.

Ritz-Laser, B., Estreicher, A., Klages, N., Saule, S., and Philippe, J. 1999. Pax-6 and Cdx-2/3 interact to activate glucagon gene expression on the G1 control element. *J Biol Chem* **274**(7): 4124-4132.

Rivera-Perez, J.A., Mager, J., and Magnuson, T. 2003. Dynamic morphogenetic events characterize the mouse visceral endoderm. *Dev Biol* **261**: 470-487.

Ryoo, H.D. and Mann, R.S. 1999. The control of trunk Hox specificity and activity by Extradenticle. *Genes Dev* **13**: 1704-1716.

Saleh, M., Huang, H., Green, N.C., and Featherstone, M.S. 2000. A conformational change in PBX1A is necessary for its nuclear localization. *Exp Cell Res* **260**(1): 105-115.

Selleri, L., Depew, M.J., Jacobs, Y., Chanda, S.K., Tsang, K.Y., Cheah, K.S., Rubenstein, J.L., O'Gorman, S., and Cleary, M.L. 2001. Requirement for Pbx1 in skeletal patterning and programming chondrocyte proliferation and differentiation. *Development* **128**(18): 3543-3557.

Shashikant, C.S., Bieberich, C.J., Belting, H.G., Wang, J.C., Borbely, M.A., and Ruddle, F.H. 1995. Regulation of Hoxc-8 during mouse embryonic development: identification and characterization of critical elements involved in early neural tube expression. *Development* **121**(12): 4339-4347.

Shen, W.F., Montgomery, J.C., Rozenfeld, S., Moskow, J.J., Lawrence, H.J., Buchberg, A.M., and Largman, C. 1997. AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins. *Mol Cell Biol* **17**(11): 6448-6458.

Shimizu, T., Bae, Y.K., Muraoka, O., and Hibi, M. 2005. Interaction of Wnt and caudal-related genes in zebrafish posterior body formation. *Dev Biol* **279**(1): 125-141.

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**(6): 1025-1036.

Subramanian, V., Meyer, B.I., and Gruss, P. 1995. Disruption of the murine homeobox gene Cdx1 affects axial skeletal identities by altering the mesodermal expression domains of Hox genes. *Cell* **83**(4): 641-653.

Suh, E., Chen, L., Taylor, J., and Traber, P.G. 1994. A homeodomain protein related to caudal regulates intestine-specific gene transcription. *Mol Cell Biol* **14**(11): 7340-7351.

Sun, X., Meyers, E.N., Lewandoski, M., and Martin, G.R. 1999. Targeted disruption of Fgf8 causes failure of cell migration in the gastrulating mouse embryo. *Genes Dev* **13**: 1834-1846.

Tabaries, S., Lapointe, J., Besch, T., Carter, M., Woollard, J., Tuggle, C.K., and Jeannotte, L. 2005. Cdx protein interaction with Hoxa5 regulatory sequences contributes to Hoxa5 regional expression along the axial skeleton. *Mol Cell Biol* **25**(4): 1389-1401.

van den Akker, E., Forlani, S., Chawengsaksophak, K., de Graaff, W., Beck, F., Meyer, B.I., and Deschamps, J. 2002. Cdx1 and Cdx2 have overlapping functions in anteroposterior patterning and posterior axis elongation. *Development* **129**(9): 2181-2193.

van Nes, J., de Graaff, W., Lebrin, F., Gerhard, M., Beck, F., and Deschamps, J. 2006. The Cdx4 mutation affects axial development and reveals an essential role of Cdx genes in the ontogenesis of the placental labyrinth in mice. *Development* **133**(3): 419-428.

Wang, P., Liu, T., Li, Z., Ma, X., and Jin, T. 2006. Redundant and synergistic effect of Cdx-2 and Brn-4 on regulating proglucagon gene expression. *Endocrinology* **147**(4): 1950-1958.

Wang, W.C., Anand, S., Powell, D.R., Pawashe, A.B., Amemiya, C.T., and Shashikant, C.S. 2004. Comparative cis-regulatory analyses identify new elements of the mouse Hoxc8 early enhancer. *J Exp Zoolog B Mol Dev Evol* **302**(5): 436-445.

Wei, C.L., Wu, Q., Vega, V.B., Chiu, K.P., Ng, P., Zhang, T., Shahab, A., Yong, H.C., Fu, Y., Weng, Z., Liu, J., Zhao, X.D., Chew, J.L., Lee, Y.L., Kuznetsov, V.A., Sung, W.K., Miller, L.D., Lim, B., Liu, E.T., Yu, Q., Ng, H.H., and Ruan, Y. 2006. A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**(1): 207-219.

Williams, T.M., Williams, M.E., and Innis, J.W. 2005a. Range of HOX/TALE superclass associations and protein domain requirements for HOXA13:MEIS interaction. *Dev Biol* **277**(2): 457-471.

Williams, T.M., Williams, M.E., Kuick, R., Misek, D., McDonagh, K., Hanash, S., and Innis, J.W. 2005b. Candidate downstream regulated genes of HOX group 13 transcription factors with and without monomeric DNA binding capability. *Dev Biol* **279**(2): 462-480.

Wolberger, C., Vershon, A.K., Liu, B., Johnson, A.D., and Pabo, C.O. 1991. Crystal structure of a MAT alpha 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell* **67**(3): 517-528.

Xu, F., Li, H., and Jin, T. 1999. Cell Type-specific Autoregulation of the Caudal-related Homeobox Gene Cdx-2/3. *J Biol Chem* **274**(48): 34310-34316.

Yamaguchi, T.P., Harpal, K., Henkemeyer, M., and Rossant, J. 1994. fgfr-1 is required for embryonic growth and mesodermal patterning during mouse gastrulation. *Genes Dev* **8**(24): 3032-3044.

Zaffran, S. and Frasch, M. 2005. The homeodomain of Tinman mediates homo- and heterodimerization of NK proteins. *Biochem Biophys Res Commun* **334**(2): 361-369.

Zakany, J., Kmita, M., Alarcon, P., de la Pompa, J.L., and Duboule, D. 2001. Localized and transient transcription of Hox genes suggests a link between patterning and the segmentation clock. *Cell* **106**(2): 207-217.

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**(5): 2920-2932.