FUNCTIONAL DISSECTION OF THE ARISTALESS-RELATED HOMEOBOX PROTEINS, ARX AND RX

DISSertation

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By

AMY N. FULLenkAMP, B.S.

*****

The Ohio State University
2008

Dissertation Committee:
Professor HEITHEM EL-HODIRI, Advisor
Professor MICHAEL OSTROWSKI
Professor DENIS GUTTRIDGE
Professor PAUL MARTIN

Approved by

Advisor
Molecular, Cellular, Developmental Biology Program
ABSTRACT

Transcriptional repression and activation are both necessary for the precise expression of the genes involved in the proper development of an organism. Aberrations in transcription factors that lead to dysfunctional proteins result in disease. Understanding the mechanisms in which transcription is regulated is vital to the identification of the factors and genetic networks that contribute to normal development. This information also provides a better understanding of the role mutations play in causing disruption in transcription factor function that leads to the impairment of developmental processes.

Homeobox genes are important for normal development. \textit{Arx} is a member of the \textit{paired-type Aristalless}-related family of homeobox proteins. Human mutations in \textit{Arx} are associated with X-linked mental retardation and other neurological disorders. Findings from the studies of \textit{Arx} deficient mice indicate the importance of \textit{Arx} on brain development and neural cell migration. Therefore, \textit{Arx} has emerged as a gene that is essential for normal brain development and is of major clinical significance. Despite what is known about \textit{Arx} clinically, less is known about its molecular functions as a transcription factor and its downstream targets. It has been shown that \textit{Arx} functions as a
transcriptional repressor, but it is not clear how disease-associated mutations affect its function as a repressor.

This dissertation describes the consequence of two disease-associated mutations on the function of Arx as a transcriptional repressor. Two independent repression domains: the octapeptide and a novel domain located in the C-terminus were identified in Arx as well as the co-factors that mediate the repression through these domains. This work also describes the DNA binding ability of Arx. Additionally, the transcriptional function of another Aristaless-related homeobox protein, Retinal Homeobox (Rx) is also reported. Rx has been shown to be an activator of transcription, however this work found that the Rx has the ability to mediate repression. Together, the goal of this dissertation is to provide insight into the transcriptional functions of two related but functionally different paired-type Aristaless-related homeobox proteins: Arx and Rx. Furthermore, this work provides evidence for the function of the octapeptide motif in the context of a transcriptional repressor and an activator.
Dedicated to my husband
ACKNOWLEDGMENTS

I would like to acknowledge and thank my advisor, Dr. Heithem El-Hodiri for giving me the opportunity to work and learn in your laboratory. I would like to thank the MCDB program for all of its support during graduate school. In addition, I would like to thank the Center for Human and Molecular Genetics for their constructive comments during presentations and for sharing reagents and protocols. I would also like to express my gratitude to the Raymond E. Mason Foundation for funding support throughout my training.

I would like to acknowledge my graduate committee: Dr. Michael Ostrowski, Dr, Denis Guttridge, Dr. Paul Martin and my past committee members: Dr. Helen Chamberlin and Dr. Aki Otoshi. Thank you for your support, guidance, and time. I have learned so much from your advice both in science and abroad.

I would like to thank my family. To my parents, for being supportive and listening even when things were difficult to explain and for ensuring me that I could and would make it. Thank you to my brother, for always being curious and conscientious about what I do and reminding me of all of the possible future opportunities. Thank you to my grandparents for always inquiring and caring about my education. Thank you to my extended family and in-laws for words of encouragement.
I would also like to thank all the members of the lab for creating a fun environment to work in and allowing me to bounce science ideas off of you. I would especially like to acknowledge Holly, Yi, and Annie for all of our discussions in the lab, outside the lab, and at lunch. I had a lot of fun talking about science, careers, and life in general. Your conversations helped bad days go faster and good days even better and for that I thank you.

Finally, I would like to thank my husband, Dave. I could not have done this without all of your words of encouragement, your support, your belief in me, your calmness (no worry attitude), your patience, and most importantly, your love. Thank you for everything and for being my source of stability throughout this entire experience.
VITA

February 1, 1979 .................................................................Born – Steubenville, OH.
2002 .................................................................................. Bachelor of Science.
                      The Ohio State University.
2002-2008 .................................................................Graduate Research Associate.
                      The Ohio State University.

PUBLICATIONS

Research Publications


FIELDS OF STUDY

Major Field: Molecular, Cellular, and Developmental Biology
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Homeobox genes and the homeodomain</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>The <em>paired</em> class of homeobox genes</td>
<td>3</td>
</tr>
<tr>
<td>1.3</td>
<td>Homeobox genes and disease</td>
<td>5</td>
</tr>
<tr>
<td>1.4</td>
<td>X-linked mental retardation (XLMR)</td>
<td>6</td>
</tr>
<tr>
<td>1.5</td>
<td>Brain development and neural migration</td>
<td>8</td>
</tr>
<tr>
<td>1.6</td>
<td>The <em>Aristaless</em> related homeobox gene</td>
<td>10</td>
</tr>
<tr>
<td>1.7</td>
<td>Importance of <em>Arx</em> in brain development</td>
<td>12</td>
</tr>
<tr>
<td>1.8</td>
<td>Functional characteristics of the Arx protein</td>
<td>15</td>
</tr>
<tr>
<td>1.9</td>
<td>Transcriptional repression: Groucho and CtBP</td>
<td>16</td>
</tr>
<tr>
<td>1.10</td>
<td>The Retinal Homebox (Rx) protein</td>
<td>19</td>
</tr>
<tr>
<td>1.11</td>
<td>Summary</td>
<td>19</td>
</tr>
<tr>
<td>2.</td>
<td>X-linked mental retardation-associated mutations affect the function of ARX as a transcriptional repressor</td>
<td>25</td>
</tr>
<tr>
<td>2.1</td>
<td>Abstract</td>
<td>25</td>
</tr>
</tbody>
</table>
2.2 Introduction

2.3 Experimental Procedures
   2.3.1 Cell Culture and transfections
   2.3.2 Plasmids and mutagenesis
   2.3.3 Immunoprecipitation
   2.3.4 RNA microinjection
   2.3.5 In situ hybridization

2.4 Results
   2.4.1 Arx is a transcriptional repressor
   2.4.2 Arx repression is mediated in part by the octapeptide motif and the conserved L33 is important for repression
   2.4.3 Two independent repression domains are responsible for repression by ARX
   2.4.4 The octapeptide mediates repression via the Groucho family of corepressors
   2.4.5 C-terminal binding protein 1 (xCtBP1) mediates repression of Arx
   2.4.6 The ORD mediates repression via CtBP
   2.4.7 Arx interacts with both Groucho and CtBP
   2.4.8 Arx contains additional conserved domains

2.5 Discussion

3. Arx Binding Site Identification
   3.1 Abstract
   3.2 Introduction
   3.3 Experimental Procedures
      3.3.1 Electrophoretic Shift Assay (EMSA)
      3.3.2 Systematic Evolution of ligands by exponential enrichment (SELEX)
   3.4 Results
      3.4.1 Arx does not bind a typical paired-type homeodomain DNA binding motif
      3.4.2 SELEX binding site identification used as an alternative approach to locate an Arx binding site
   3.5 Discussion

4. Rx N-terminal Functional Dissection
   4.1 Abstract

ix
4.2 Introduction...........................................................................................................78
4.3 Experimental Procedure.......................................................................................80
  4.3.1 Cell culture and transfections.............................................................80
  4.3.2 Plasmids and mutagenesis.................................................................81
4.4 Results..................................................................................................................81
  4.4.1 Rx is a weak activator despite the presence of an octapeptide motif........82
  4.4.2 N-terminal mutations and deletions affect the function of Rx as a weak activator .................................................................................................83
  4.4.3 Amino acids 1-9 of the N-terminus do not affect Rx function ...94
4.5 Discussion............................................................................................................85

5. Discussion...............................................................................................................96

  5.1 Introduction.......................................................................................................96
  5.2 Results...............................................................................................................97
    5.2.1 Chapter 2...............................................................................................97
    5.2.2 Chapter 3...............................................................................................98
    5.2.3 Chapter 4...............................................................................................98
  5.3 Future Directions ..............................................................................................99
    5.3.1 Chapter 2...............................................................................................99
    5.3.2 Chapter 3.............................................................................................106
    5.3.3 Chapter 4.............................................................................................109
  5.4 Concluding Remarks.........................................................................................111

Bibliography .............................................................................................................113

x
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Mouse brain region schematic</td>
<td>20</td>
</tr>
<tr>
<td>1.2 Schematic of neuronal migration</td>
<td>21</td>
</tr>
<tr>
<td>1.3 Schematic of known Arx mutations to date</td>
<td>22</td>
</tr>
<tr>
<td>1.4 Schematic of Arx protein conservation and domains</td>
<td>23</td>
</tr>
<tr>
<td>2.1 <em>Xenopus</em> Arx functions as a transcriptional repressor</td>
<td>41</td>
</tr>
<tr>
<td>2.2 Arx mutation constructs</td>
<td>42</td>
</tr>
<tr>
<td>2.3 The octapeptide domain is necessary for repression</td>
<td>43</td>
</tr>
<tr>
<td>2.4 Arx mutation constructs</td>
<td>44</td>
</tr>
<tr>
<td>2.5 Two repression domains are responsible for repression by Arx</td>
<td>45</td>
</tr>
<tr>
<td>2.6 Whole mount in situ hybridization of xEsg-1</td>
<td>46</td>
</tr>
<tr>
<td>2.7 The octapeptide mediates repression via the Groucho family of co-repressors</td>
<td>47</td>
</tr>
<tr>
<td>2.8 Double whole mount in situ hybridization of xArx and xCtBP1</td>
<td>48</td>
</tr>
<tr>
<td>2.9 C-terminal binding protein (CtBP1) mediates repression of Arx</td>
<td>49</td>
</tr>
<tr>
<td>2.10 CtBP enhances repression in constructs with an intact OP</td>
<td>50</td>
</tr>
<tr>
<td>2.11 The ORD mediates repression by CtBP1</td>
<td>51</td>
</tr>
</tbody>
</table>
2.12 Deletion of the PLGLS motif does not disrupt repression by CtBP ........52
2.13 Arx interacts with Groucho and CtBP .........................................53
2.14 Arx contains additional conserved regions ........................................54
2.15 Arx contains two independent repression domains .............................55
3.1 EMSA probes ....................................................................................72
3.2 AI/CI EMSA ........................................................................................73
3.3 P3 EMSA ............................................................................................74
3.4 Consensus sequences obtained from SELEX ........................................75
3.5 Homeodomain Schematic ......................................................................76
4.1 Rx is a weak activator ..........................................................................89
4.2 Octapeptide Comparison .......................................................................90
4.3 Rx constructs ........................................................................................91
4.4 Rx exhibits greater activity upon deletion or mutation of the OP ..........92
4.5 The N-terminus of Rx affects its function ............................................93
4.6 Rx constructs ........................................................................................94
4.7 Amino acids 1-9 in the N-terminus do not affect Rx function ...............95
CHAPTER 1

INTRODUCTION

1.1 Homeobox genes and the homeodomain

Homeotic genes of *Drosophila* are involved in the specification of the developmental pathways that give rise to the antero-posterior body axis segments of the fly with each segment having its own unique identity [1]. Mutations in homeotic genes result in the misexpression of body structures in different segments of the fly, demonstrating that they are responsible for specifying the body segments [1]. These genes often exist in clusters in the *Drosophila* genome; two such clusters are the bithorax complex and the Antennapedia complex both located on chromosome three [2]. The Antennapedia complex was found to control segmental development in the head and thorax while the bithorax complex controls more posterior segments of the fly [3-6]. It was hypothesized that the proximity of these gene clusters were indicative of the genes evolving by mutational diversification of a common ancestral gene or complex [6, 7]. It was later found that these clusters of genes contained a conserved element that was identified to be a region preserved by selection against mutational change [1, 2]. The conserved element was named the homeobox and was used to identify other genes with
this sequence homology to be included in the homeotic gene family. The homeobox is a 180 base pair DNA element discovered in *Drosophila* that has been shown to occur in homeotic genes of all metazoa including sponges, plants, fungi, vertebrates and humans [8]. The homeobox sequence encodes a 60 amino acid DNA binding domain which represents the homeodomain that consists of three \( \alpha \)-helices folded around a hydrophobic core and a flexible N-terminal arm that becomes ordered upon binding to DNA [9]. Homeodomain proteins primarily serve as transcription factors that regulate target genes in a precise spatial and temporal pattern and the homeodomain allows for the sequence-specific recognition of target genes [8]. X-ray crystallographic and NMR spectroscopic experiments on several members of the homeodomain protein family have shown that helix I and helix II lie parallel to each other and helix III lies across from them. It is helix III that has been shown to bind to the major groove of DNA where it makes specific contacts with the DNA and so is known as the recognition helix [8]. Furthermore, it has been found that despite the multiple processes in which homeodomain proteins are involved, the majority that have been characterized all recognize DNA sequences containing a 5’-TAAT-3’ core motif [8]. In fact the residues that have major implications on DNA binding: Tryptophan (Trp) 48, Phenylalanine (Phe) 49, Asparagine (Asn) 51, and Arginine (Arg) 53 are within the recognition helix and are all absolutely conserved in homeodomain proteins [10]. Residues that contact the bases of the TAAT motif have been identified to be positions 2, 3, and 5 in the N-terminal arm and positions 47 and 51 of the recognition helix within the homeodomain [11]. Alignment based on the three-dimensional structure of the homeodomain have allowed for the segregation of the
proteins into six distinct classes that is consistent with the known functional and structural characteristics of the proteins [10]. The six classes are: HOX, NK, Paired, LIM, POU, and atypical. This dissertation focuses on members of the paired class of homeodomain proteins.

1.2 The paired class of homeobox genes

The paired class (Prd-class) of homeodomain proteins is a major class of homeobox genes and is involved in fundamental developmental functions. Several genes of this family have been identified in cnidarians indicating that this class emerged early in evolution [12]. The paired class proteins are defined by the presence of a homeodomain that resembles that of the Drosophila paired protein. Three sub-classes have been identified for the paired class based on the residue at position 50 within the homeodomain, which is a key residue involved in DNA binding specificity [13]. The Pax- or prd-type subgroup contains a serine residue at position 50 of the homeodomain (S50) and contains a second DNA-binding domain called the prd domain [14, 15]. Another subgroup prd-like, related to the Orthodenticle gene, contains a lysine residue at position 50 (K50) of the homeodomain and lacks a prd domain [16]. The third subgroup also called prd-like contains a glutamine at position 50 of the homeodomain and also lacks a prd domain [17-19]. One specific subset of the paired class, the Aristaless-related proteins, contains a glutamine at position 50 (Q50) of the homeodomain and has a conserved domain of unknown function located at the C-terminus called the Aristaless (or OAR) domain [13]. The Drosophila Aristaless (al) gene was the first to be described for this group of proteins. Mutations in the al gene resulted in a phenotype restricted to a
limited area of the fly, the aristae (a subset of appendages), which were reduced in size
depending on the mutation [18].

The paired type homeodomain proteins have been shown to dimerize to mediate
cooperative binding on sites that contain two palindromic TAAT sequences that are
separated by two or three less significant nucleotides (P2 or P3 sites) [20]. In addition, it
has been found that the Pax-type (S50) homeodomains prefer to bind to P2 sites, while
other paired type homeodomains (Q50 and K50) including those of the Aristalless-related
homeodomains prefer to bind to P3 sites (TAATNNNATTA) [13]. The Q50
homeodomains have also been shown to prefer YNR sequences in the middle of the
TAAT sequences (Y=C or T; R= A or G) [13]. Furthermore, Wilson et al. [20] reported
that Prd class homeodomains contain sufficient information to specify the cooperative
binding as homo- or heterodimers on sequences composed of two TAAT half sites. This
cooporative binding is thought to occur because the first binding event to one of the half
sites, may change the conformation of the second half site, which could improve the
affinity for another homeodomain to bind [20]. It was also found that the ninth amino
acid of the recognition helix has the ability to perform three functions [20]. The three
functions include: the ability to specify the spacing between the TAAT half sites, to
modulate the cooperative interactions, and to determine the base pairs in the variable part
of the palindromic sequence [11, 20]. Therefore, the paired class homeodomain proteins
harbor the ability to heterodimerize with other homeodomain proteins allowing for
binding specificity despite the inherent ability of homeodomains to bind similar DNA
sequences.
1.3 **Homeobox genes and disease**

Homeobox genes are involved in many important cellular processes such as proliferation, differentiation, and migration. In addition, homeobox genes are also involved in embryonic development in mechanisms such as limb formation, axial skeleton patterning, craniofacial morphogenesis, central nervous system development, and organogenesis [21]. Considering the multiple processes that involve homeobox genes, it is not surprising that disruption of these genes would lead to dysfunction in the pathways they are part of and result in disease. Among the disorders that homeobox genes are known to be involved in are: cancer, heart disease, and neurological disorders.

It has been reported that many homeobox genes contribute to cancer development in several ways such as deregulation of their function through chromosomal translocation, deletion of genes via loss of heterozygosity, and by epigenetic mechanisms that deregulate gene expression [21]. This is understandable when considering that the molecular events that take place during embryogenesis are similar to the events that occur during tumorigenesis and that homeobox proteins are involved in cell cycle control [22].

It has also been shown that the Nkx class of homeobox genes is critical for normal cardiac development and mutations in these genes are associated with congenital heart disease [23]. Recently, there has been interest in the involvement of homeobox genes in neurological disorders. A number of homeobox genes have been shown to be involved in neural development and aberrations in these genes result in diseases such as epilepsy and mental retardation [24, 25].
1.4 X-linked mental retardation (XLMR)

Mental Retardation (MR) is a disability characterized by significant limitations in both intellectual function and adaptive behavior in areas such as conceptual, social, and practical skills that manifests before the age of eighteen [26, 27]. Persons affected with this disability have a reduced capacity to function on a cognitive and social level and therefore a better understanding of etiology of this disorder is of great importance to human health and society. MR is one of the main causes a child is referred by their pediatricians to neurologists or geneticists and is one of the most challenging diagnoses to access because it is a very heterogeneous disorder. It is estimated that MR affects 1-3% of the population [28-30]. MR is a disability that can be caused by several genetic and environmental factors. Such factors can act on the development of the brain and central nervous system at critical stages of ones life including prenatally, perinatally, and postnatally due to brain ischemia, fetal alcohol syndrome, or infection [30-32]. The genetic factors that contribute to MR include: large chromosome abnormalities such as trisomy 13 in Down’s syndrome, microdeletions in which a very small portion of the genome is deleted, copy-number variation and gene duplications, or coding abnormalities in a single gene (monogenic) [33].

MR has been found predominately in males with a ratio of 1: ~1.3 as compared to females [33] and averages to about 30% male excess worldwide [34]. The higher incidence in males was initially thought to be due to social biases, however, it had been postulated that the increase was due to genes on the X-chromosome. The identification of many families in which MR segregated in an X-linked inheritance fashion and
epidemiologic studies confirmed this notion. Furthermore, the discovery of a fragile site at Xq27.3 that segregated with MR in males was the first mapping for the Fragile X syndrome [34, 35]. The Fragile X gene was later found by the identification of an unstable trinucleotide repeat that segregated with the syndrome and the gene responsible, FMR1 was subsequently cloned [36-39]. Since then, Fragile X syndrome has been the most common form of X-linked mental retardation (XLMR) and been studied extensively [33, 40]. In fact, the research involved in Fragile X syndrome has provided insights into the molecular mechanisms of brain function and disease. In 2003, a mouse deficient in the FMR1 gene was found to resemble phenotypic aspects of Fragile X syndrome [41, 42]. More recently, a Drosophila model for Fragile X syndrome was developed in which the FMR1 homolog, dfmr1 was deficient [43]. It was found that these flies had neuronal and behavioral phenotypes similar to the symptoms observed in humans with Fragile X and that treatment with mGluR (metabotropic glutamate receptor) was able to restore these abnormalities [43]. This finding suggests that a similar treatment may be efficient for treating humans to aid in restoring cognitive defects and provides hope for finding treatments for MR.

Families in which no clinical features outside of MR were observed in which Fragile X syndrome was not the cause, led to the term non-specific XLMR [33, 44]. XLMR has been divided into two sub-types: syndromic and non-syndromic. The syndromic form is associated with symptoms other than MR including a specific pattern of physical, neurological, or metabolic abnormalities are present upon physical examination, laboratory investigation and brain imaging [45, 46]. The non-syndromic
form is associated with no other consistent phenotypes other than MR [44]. A recent XLMR update has reported that XLMR is a common cause of inherited intellectual disability with an estimated prevalence of ~1/1000 males [31]. Of the 1200 genes on the X chromosome it has been speculated that mutations within 200 or more of them may be associated with MR [47]. This is of even more interest when considering that 40% of these are genes are expressed in the brain [40]. In 2008, it was reported that there are 82 XLMR genes that are associated with 215 conditions in which 149 were syndromic and 66 non-syndromic forms of XLMR [31].

1.5 Brain development and neural migration

The human brain forms during the third week of gestation from the rostral end of the neural tube [48]. The central nervous system is a complex structure and the orchestration of crucial cellular events is necessary to ensure its proper organization. The cerebral cortex is the outer layer of the brain and is a laminated structure composed of six layers of neurons, nerve fibers, and supporting cells [48]. The cerebral cortex is responsible for sensory, motor, and association functions of the brain and connections are made to this region from other areas of the brain to transmit information for these processes. The surface of the cortex contains many folds (gyri) and fissures (sulci) that are the result of the migration of neurons. This feature is due to neurons that are generated in the germinal zone lining the ventricular zone (VZ) that must migrate to their final positions where they organize and interconnect to form the cerebral cortex [48]. These folds in the brain create more surface area and allow for an increase in the amount of information that can be processed. Through a mechanism that involves mutually
repressive interactions, transcription factors establish the boundaries and patterning of the brain before the cells migrate to their final destinations [49]. Disruption of this process results in a smooth cerebral surface or lissencephaly, which is associated with many forms of mental retardation and epilepsy [50].

The most characterized mode of neural migration is radial cell migration in which cells that are born in the ventricular zone migrate radially along specialized glial fibers to their final location [51]. Another mode of migration, tangential or non-radial migration, in which cells born in the ventricular zone migrate along pathways perpendicular to radial migration has also been described [51]. Recently, studies have been focused on tangential migration of neurons in the telencephalon or forebrain. The telencephalon contains two major regions: the pallium or roof and the subpallium or base. The pallium gives rise to the cerebral cortex and hippocampus and the subpallium contains three subdivisions: the striatal, pallidal and telencephalic stalk domains (Fig.1.1) [49]. It has been shown that cells derived from the subpallial telencephalon contribute to the population of tangentially migrating cells in the cortex [52]. Regions of the subpallial telencephalon responsible for the production of cells that migrate in tangential streams are the lateral ganglionic eminence (LGE) and the medial ganglionic eminence (MGE) (Fig.1.2) [52]. Furthermore, the GABA (γ-aminobutyric acid)-producing interneurons have been identified as cells that are born in the subpallial telencephalon and reach the cortex by tangential migration [52]. It is thought that tangential migration occurs when cells that are specified in dorsal regions of the telencephalon are required in ventral
structures and that this may be a mechanism selected through evolution to increase the
cellular complexity of neuronal circuits [49, 52].

Genes involved in neuronal migration have been identified such as LIS1, Filamin
1, DCX, RELN, and ARX [51]. Many of these genes are located on the X-chromosome
and result in lissencephaly when mutated [50]. The mechanism by which these genes
result in disease not completely understood, however some of these genes are involved in
cell morphology and motility as a result of their role in the regulation of the actin
cytoskeleton [51]. Further studies are necessary to unravel the biology of these genes in
the important process of cellular migration in the brain.

1.6 The Aristalless related homeobox gene

The Aristalless related homeobox gene (Arx) located on Xp22.11 is mutated in
patients with syndromic or non-syndromic forms of XLMR [53-55]. Arx is a member of
the paired type homeobox genes and belongs to the Aristalless-like subgroup of genes that
are predominantly expressed in the central and/or peripheral nervous system [13]. Arx is
expressed in the telencephalon, diencephalon, in the midbrain/hindbrain border, floor
plate of the spinal cord, somites, and testis in mouse and zebrafish where it was
speculated to be involved in axonal guidance of the floor plate [56]. In addition, another
group described expression of Arx in the mouse pancreas [57]. Recently, marsupial
orthologues of MR genes (which diverged from humans about 180 million years ago)
were studied to determine whether these genes on the X chromosome were recruited to
the X chromosome or whether their brain-specific functions were acquired after
relocation to the mammalian X chromosome [58]. It was found that of all the genes they
studied, Arx was the only one found to have retained its brain-specific expression for over 180 million years suggesting its fundamental role in mammalian brain function and development [58].

Mutations in Arx cause X-linked West syndrome (infantile spasms, hypsarrhythmia and MR), X-linked myoclonic epilepsy with spasticity and intellectual disability, Partington syndrome (mental retardation, ataxia, and dystonia), Proud syndrome (XLMR with agenesis of the corpus callosum and abnormal genitalia), as well as non-syndromic XLMR and autism [54, 55, 59, 60]. Arx is also mutated in patients with X linked lissencephaly with abnormal genitalia (XLAG) [53]. Lissencephaly is a brain malformation in which there is an incomplete development of the folds (gyri) of the outer region of the brain (cerebral cortex) that causes the surface of the brain to be thickened and smooth [32] and is caused by a defect in neuronal migration [61]. XLAG was first described in five unrelated children who were all males and suffered from lissencephaly, agenesis of the corpus callosum (the structure that connects the two hemispheres of the brain is absent), epilepsy, temperature dysregulation, midbrain malformations, underdeveloped genitalia, and a severely shortened lifespan [48, 59]. To date there are at least 59 mutations in Arx that are associated with seven XLMR disorders [24] (Fig.1.3). Mutations in Arx are considered to be the second most common cause of XLMR after the FMRI gene that is involved in Fragile X syndrome [40]. In fact, it has been suggested that Arx mutation screening be considered when a patient has XLMR that is different from Fragile X syndrome [24, 62]. Mutations in Arx have consistent genotype-phenotype correlation as well as remarkable pleiotropy [63]. This phenomenon
has been further defined as premature termination mutations consisting of large deletions, frameshifts, nonsense mutations, and splice-site mutations result in XLAG while missense mutations, polyalanine expansions or duplications result in less severe phenotypes [63].

1.7 Importance of Arx in brain development

Arx expression is first detected in the mouse at E8 (embryonic day 8) in a restricted region of the neuroepithelium that corresponds to the forebrain [54, 56]. Arx is expressed throughout the telencephalic structures such as the ganglionic eminences, cerebral cortex, and the hippocampus at later stages with weaker expression in the adult [54, 56]. At higher magnifications, Arx was found to be expressed in the ventricular zone of the developing cortex, which contains proliferating cells, in the intermediate zone, which contains migrating neurons, and in the developing cortical plate, which contains maturing neuronal cells [54]. It was this restricted expression pattern together with the lack of null or complete loss of function mutations found in XLMR patients that suggest that Arx is a critical factor required early for normal brain development. Reports of the expression pattern and the description of human mutations led several groups to study the effects of the loss of function of Arx in the mouse.

The first report of a mouse knockout, Kitamura et al [53] found that male mice deficient for Arx were born at E19.5, died within a half day, and had smaller brains and testes. In addition, 5-bromo-2-deoxyuridine (BrdU) injection and apoptosis experiments revealed that the thin embryonic neocortex in the mice was due to decreased proliferation of neuroepithelial cells in the ventricular zone and not because of apoptosis at E14.5 [53].
It was also found that Arx deficiency resulted in the loss of direct migration of GABAergic interneurons from the medial ganglionic eminence to the cortical intermediate zone indicating that Arx plays a role in the normal tangential migration of these neurons [53]. Moreover, these mice exhibited features to that of children with XLAG including absence of interneurons in the neocortex, a simplified three-layer cortex and other structures of the telencephalon were small or absent. Lissencephaly could not be readily assessed, as the mouse brain is naturally smooth. Taken together, this report describes how loss of function of Arx leads to deficient proliferation and incomplete migration of neurons that could be the cause of the brain malformations observed in XLAG [53].

Another group characterized the brain malformations found in the Arx deficient mice even further to address how the aberrations observed would lead to neurological disorders. Colombo et al [64] reported that Arx inactivation led to the inability of cells in the ventral forebrain to migrate radially as well as tangentially. This defect impacts the ability of these cells to correctly mature and differentiate in the basal ganglia structures [64]. In addition, a role for Arx in the regulation of cholinergic neural differentiation was also found. Because the Arx deficient mice share similar characteristics to XLAG patients, this work illustrates the potential pathophysiology of how these brain defects lead to disease.

It has been shown that Arx plays a role in the tangential migration of a subset of GABAergic interneurons [53, 65, 66]. It is thought that this is how mutations in Arx result in epilepsy because GABA is the principal inhibitory neurotransmitter in the
cerebral cortex that maintains the signal that counterbalances neuronal excitation in humans [66]. This defect in migration, detected in the mouse brain, has recently been identified in the human brain of XLAG patients [67]. This result further demonstrates the role of ARX in brain development and how it aberration results in disease.

An additional study using RNA interference (RNAi) technology or overexpression was recently done to further investigate the cell-autonomous role of Arx on corticogenesis. It was found that targeted inhibition of ARX in progenitor cells of the cerebral cortex results in their premature exit from the cell cycle and differentiation into neurons and that overexpression of ARX resulted in an increased length of the cell cycle indicating the role of ARX in maintaining progenitor proliferation [68]. Friocourt et al also described a role for ARX in the cell morphology and radial migration of pyramidal neurons. This study also provided further evidence to support the finding of the involvement of ARX on the tangential migration of interneurons from the ventral telencephalon [69]. This report emphasizes the importance of ARX in brain development and suggests that because both overexpression and inactivation of this gene affect neuronal cell proliferation and migration, a precise balance ARX expression is necessary for proper development of the cortex. In addition, this report together with similar findings from Colombo et al, provide new insight into the possible role of ARX in cell morphology and actin cytoskeleton regulation as a mechanism that when disrupted, leads to disease.

It has also been reported that Arx is expressed in the developing mouse pancreas [57]. An Arx deficient mouse was developed in which defects in the pancreas were
extensively studied, however brain abnormalities were not reported [57]. It was found that \textit{Arx} promotes the $\alpha$-cell fate of endocrine progenitor cells and does this by antagonizing the $\beta$-cell and $\delta$-cell commitment and was necessary for appropriate $\delta$-cell location [57]. It was also speculated that \textit{Arx} and \textit{Pax4} play opposing roles in endocrine cell fate determination because mice deficient in one or the other of these genes had opposite effects on pancreas cell development [57]. This was later confirmed by studies done in mice deficient for both \textit{Arx} and \textit{Pax4} in which it was shown that \textit{Arx} was required for $\alpha$-cell fate and \textit{Pax4} was necessary for $\beta$-cell destiny [70]. In addition, Collumbat \textit{et al} showed that this phenomenon was due to the ability of each factor to inhibit the transcription of each other.

\section*{1.8 Functional characteristics of the \textit{Arx} protein}

The \textit{Arx} homeobox gene encodes a protein that contains three highly conserved domains (Fig.1.4). The protein was shown to contain a homeodomain, octapeptide domain, and an \textit{Aristaless} (OAR) domain [54-56]. Comparison of ARX to its vertebrate homologs shows that it is 94.3\% identical to mouse, 57.2\% identical to zebrafish, and 59\% identical to \textit{Xenopus} [55, 71]. The octapeptide, nuclear localization sequence, and homeodomain are identical while the \textit{Aristaless} domain is identical in human and mouse and 87\% identical between human and zebrafish [55]. The high conservation of these domains suggests their functional importance. The homeodomain has been well characterized and known to be involved in binding to DNA. The octapeptide is a motif that has been conserved throughout evolution and is involved in mediating repression [72]. The \textit{aristaless} or OAR (Orthopedia, Aristaless, Rx) domain is highly conserved
among the *Aristaless*-related homeobox proteins, however its function is not known but has been suggested to play a role in activation [13, 73]. The ARX protein contains additional regions of conservation outside of the three described such as polyalanine tracts (Fig.1.4). Polyalanine (polyA) tracts have been described in several human diseases, however their function on transcription or protein-protein interactions remains to be determined [74]. Mutation in the ARX polyA tracts have been identified and are thought to cause protein accumulation and nuclear inclusions that may result in disease [75]. Several studies have linked ARX to transcriptional repression as well as a potential role in activation [76, 77]. Additional studies are necessary to elucidate how mutations in *Arx* result in aberrant function that leads to disease.

1.9 **Transcriptional repression: Groucho and CtBP**

Transcriptional repression is an important cellular event that has become a topic of interest in understanding the molecular mechanisms involved in regulation of gene expression during developmental processes. There are two general modes of repression: long-range and short-range. Long-range repression involves a repressor that has the ability to render a promoter resistant to the influence of enhancers including those that are located thousands of base pairs away from the binding site of the repressor [78]. On the other hand, short-range repression acts in a local manner by blocking the function of nearby DNA-bound activators and does not interfere with those farther away [78]. Two proteins known to able to achieve these two types of repression are: Groucho and CtBP. These proteins act as co-repressors because they do not bind to DNA itself, rather they
bind proteins via physical interaction to transcriptional repressors that are already DNA-bound.

The Groucho (Gro) family of co-repressors was first described in *Drosophila* and orthologs are found in all metazoans [79]. In humans, Groucho proteins are called transducin-like enhancer of split (TLE) proteins. The Grouchos proteins contain a conserved N-terminal tetramerization domain, a variable central region, and a conserved C-terminal WD-repeat domain [78]. WD repeats consist of 40 amino acid motifs that occur in repeats and mediate protein-protein interactions [78, 80]. The Groucho co-repressor has been found to bind to transcriptional repressors through several motifs such as the WRPW and WRPY tetrapeptide motif as well as through the a short peptide motif called the Engrailed homology-1 (Eh1) or octapeptide motif [78, 81]. It is thought that Groucho proteins form complexes that bring other factors that mediate repression to the DNA such as other co-repressors and chromatin remodeling factors [78, 82].

The function of the Groucho co-repressor can be modulated in different ways. One ways is through post-translational modification such as phosphorylation. Groucho can be phosphorylated by the homeodomain-interacting protein kinase 2 (HIPK2). This phosphorylation results in Groucho dissociation from the co-repressor complex, which relieves the transcriptional repression of its downstream target genes [83]. Therefore, phosphorylation of Groucho may be an important modification in the regulation of repression of genes involved in developmental processes. Another way in which Groucho function is regulated is through the interaction with other Groucho proteins. Two shorter forms of Groucho, Grg5 and Grg6, are able to bind to the full-length form of
Groucho and antagonize its repressive activity [81]. It is thought that in certain cellular contexts, the shorter forms of Groucho can act as dominant-negatives to inhibit the repressive function of Groucho.

C-terminal binding proteins (CtBP) were first discovered as a protein that interacted with the C-terminus of the adenovirus E1A protein [84-86]. CtBP gene products are found in higher eukaryotes and are essential for normal animal development [84]. There are two CtBP genes found in vertebrates (CtBP 1 and 2), however CtBP1 is known for its activity as a transcriptional co-repressor [84]. To date, more than 30 different transcription factors have been reported to modulate their activity by recruitment of CtBP [84]. It has been shown CtBP can bind transcriptional repressors through a (PXDLS) binding motif [87, 88]. CtBP is thought to primarily exert its co-repressive functions by binding to chromatin remodeling factors such as histone deacetylases (HDACs), however recently it has been shown that CtBP mediate repression through its modification by SUMOylation [78, 84].

Groucho and CtBP mediate different modes of repression and both co-repressors are required for proper vertebrate development. The multi-functionality of co-repressors is speculated to allow for the specific deregulation of gene expression in a way that is tailored to the goal of the repression [78]. The presence of both types of co-repressors indicates that multiple repression mechanisms may be necessary for different developmental processes. Recently, many reports have shown that some transcriptional repressor proteins can bind to both Groucho and CtBP to modulate their repressive functions [84] indicating that these two factors may work together in certain situations.
1.10 The Retinal Homeobox (Rx) Protein

The retinal homeobox gene (Rx) is a member of the paired-type homeobox family and is included in the same subfamily as Arx. Rx is expressed in the developing retina of vertebrates and is involved in the earliest steps of eye development [13]. The Rx protein contains the characteristic protein domains found in this family: an octapeptide, homeodomain, and OAR domain. Similar to Arx, little is known about the functional significance of these domains on transcription. This work describes the functional relevance of the Rx octapeptide on transcription and how its activity may be modulated in comparison to the Arx octapeptide. Furthermore, this dissertation demonstrates how the octapeptide may function in the context of two related, yet functionally different homeobox proteins.

1.11 Summary

Transcription factors regulate gene expression through both activation and repression of downstream genes. Mutations in transcription factors that alter their functions result in the deregulation of the genetic networks they are involved and result in disease. A multitude of neurological disorders result from mutations in Arx. It is clear from the aforementioned studies together with the clinical findings that Arx plays a crucial role in the normal development of the brain. Future experiments aimed at identifying the molecular pathways in which Arx is involved will provide a better understanding of how mutation in this gene causes such a wide variety of neurological diseases. Additionally, functional dissection of the protein will yield information on the transcriptional abilities of Arx and how it regulates its downstream targets. Identification
of potential modifier genes will also help in the understanding of how one gene contributes to many disorders.

The goal of this work was to determine the effects that disease-associated mutations have on the transcriptional functions of Arx, to identify potential downstream targets of Arx, and to study transcriptional repression of the retinal homeobox gene (Rx). This dissertation dedicates a chapter to each of these areas. Chapter 2 describes the functional consequence of disease-associated mutations on the transcriptional repression activity of Arx. In addition, Chapter 2 describes the presence of two independently regulated repression domains and provides insight into the purpose of such domains within Arx. Chapter 3 deals with the examination of possible downstream binding preferences of Arx and provides information into the types sites that Arx may bind. Finally, Chapter 3 demonstrates the potential of Rx, a weak activator of transcription, to possess a repressive function. Specifically, Chapter 3 describes the function of the Rx octapeptide and how its repressive function differs from that of the Arx octapeptide due to the presence of a different protein sequence in the N-terminus of the two proteins. Collectively, this dissertation demonstrates and compares the functional characteristics of two homeobox proteins that are essential transcriptional regulators of neural development.
Figure 1.1. Mouse brain region schematic (adapted from Rubenstein et al) [49]. A. Schematic of a sagittal section of an E12.5 mouse brain. B. Schematic of a transverse section of an E12.5 mouse brain. MGE=medial ganglionic eminence, LGE=Lateral ganglionic eminence.
Figure 1.2. Schematic of neuronal migration (adapted from McManus et al) [51]. A. Schematic diagram of a coronal section of the forebrain indicating radially and tangentially migrating neurons. Cp: cortical plate, iz: intermediate zone, vz: ventricular zone
Figure 1.3. Schematic of known Arx mutations to date (adapted from Gecz et al) [24]. A. Schematic showing the type and location of the 68 ARX mutations known to date.
Figure 1.4. Schematic of Arx protein conservation and domains (adapted from El-Hodiri et al. [71]). A. Conservation between human, mouse, Xenopus, and zebrafish. Black regions are completely conserved. Underline portion indicates the characteristic amino acids of that domain. P(A)=polyalanine tract.
2.1 Abstract

The *Aristaless* related homeobox (Arx) is mutated in patients with X-linked mental retardation and a range of other neurological diseases. Arx is a *paired*-type homeobox protein that contains three highly conserved protein motifs: the octapeptide, homeodomain, and OAR domain. Many disease-causing mutations occur within the conserved regions of the protein; however, the molecular consequences of the mutations are unclear. Here, we show that two disease-associated mutations disrupt the function of Arx as a transcriptional repressor. We found that Arx contains two independent repression domains: the octapeptide motif that is regulated by the co-repressor Groucho and a novel domain located in the C-terminus that is regulated by CtBP.
2.2 Introduction

The *Aristaless* related homeobox gene, Arx, has come to the forefront of studies involving brain development and has been recently described as one of the most important disease causing genes on the X chromosome [24]. Mutations in ARX have been implicated in a multitude of human neurological disorders including X-linked lissencephaly, X-linked infantile spasms or West syndrome, Partington syndrome, Proud syndrome, X-linked lissencephaly with abnormal genitalia (XLAG) and non-syndromic mental retardation [68]. Loss of function studies in mice have shown that Arx is involved in promoting neuronal proliferation, differentiation, and the proper migration of GABA-ergic interneurons, further indicating the importance of this gene in brain development [53, 64, 66].

Arx is a *paired*-type homeodomain containing transcription factor that is a regulator for essential developmental events in vertebrates. The protein shares high sequence conservation with the *Drosophila aristaless* (al) protein especially within the octapeptide, homeodomain, and OAR (*Aristaless*) motifs [13, 55]. The octapeptide motif or engrailed homology motif (eh-1) is involved in repression [24], the homeodomain confers DNA binding, and the function of the OAR motif is not known. The high conservation of Arx suggests that the conserved regions must be functionally important and that mutation in these domains would lead to the disruption of the protein and the transcriptional pathways in which it is involved. Indeed, there are several ARX
mutations located at key residues within the conserved motifs that are speculated to disrupt the function of the protein [24].

Several studies, both in vitro and in vivo have shown that Arx is a transcriptional repressor [70, 76, 77]. Molecular analysis to determine the regions that are essential for repression would provide insight into how mutations in these regions cause aberrant function and therefore, disease.

In this work, we focused on elucidating the regions of Arx that are important for repression and the co-factors that mediate repression. Through molecular dissection of the protein, utilizing disease-associated mutations as a guide, we identified two independently regulated repression domains within conserved regions of Arx. We also identify a novel interaction between Arx and the co-repressors Groucho and CtBP.

2.3 Experimental Procedures

2.3.1 Cell Culture and Transfections

Cos7 cells were maintained in Dulbecco’s modified Eagle medium that was supplemented with 10% Calf Bovine Serum, 100 units of the antibiotics penicillin/streptomycin, and 2mM L-glutamine at 37°C at 5% CO₂. Sub-confluent cells were trypsinized and plated the day before transfection at a density of 5X10⁴ cells per well in a 24-well plate. Cells were transfected in triplicate with Fugene 6 transfection reagent at a 3:1 DNA to Fugene ratio per manufacturer’s protocol (Roche). Empty pCS2-Gal4 plasmid was added to bring all DNA totals to 400ng per reaction. A Renilla luciferase reporter construct under the control of the thymidine kinase promoter (pRL-TK, Promega) was co-transfected as an internal control. The internal control was
transfected at a 1:20 ratio to the Gal4 responsive luciferase reporter (Gal4-Tk-luc) plasmid together with either no additional DNA, (the transcriptional baseline control), or DNA from wildtype or mutated constructs. Cells were harvested 24 hrs later and lysates were assayed for firefly and Renilla luciferase reporter activities using the Dual Reporter Luciferase Assay Kit (Promega). The reporter assays were carried out as described per manufacturer’s protocol.

2.3.2 Plasmids and Mutagenesis

Transfected DNA included pCS2+Gal4 DBD plasmids containing wildtype and mutant versions of the Xenopus Arx [71]. xArx was isolated from a pBS/xArx plasmid by PCR with the primers listed below to add the proper restriction sites for cloning in frame with the Gal4 tag. The xArx PCR product was then cloned into the pCR 2.1-TOPO vector (Invitrogen), the sequence verified, and then sub-cloned into the pCS2+Gal4 vector using EcoRI and SnaBI. The pCS2+Gal4DBD vector was made by isolating the Gal4DBD from the pGBT9 plasmid using Hind III (filled in) and EcoRI restriction sites and then sub-cloned into pCS2 using BamHI (filled in) and EcoRI.

Forward: 5’GATCGAATTTCAGCAGCCACTACCAACAGGAG 3’
Reverse: 3’GATCGGATCCTCAGCAAACCTCTTTCCCTC 5’

The L33P mutation was made using the Quick Change XL Mutagenesis Kit (Stratagene) with primers designed to make the desired mutation using pCS2+Gal4 DBD/Arx as a template.

L33P Forward: 5’ TGTATAGACAGTATCCCGGGCAAGAAG 3’
L33P Reverse: 3’ CTTCTTCCGGGATACTGTCTATACA 5’
The Exon 5 truncation was made using PCR using primers that were designed to truncate a specific region and cloned into pCR 2.1-TOPO vector and then sub-cloned into the Gal4 plasmid using EcoRI.

Exon 5 Forward: 5’ GATCGAATTCAGCAGCCACTACAAACAGGAG 3’
Exon 5 Reverse: 3’ GATCCTCGAGTCTGCCGAAGGCCGGGCT 5’

Double mutation constructs were made using site-directed mutagenesis in which the plasmid containing the L33P mutation was subjected to PCR using the above primers to make the desired second mutation. The 398 truncation construct was made using a NotI site within the Gal4-Arx construct to remove the sequence following the site and re-ligate the remaining fragment.

The xGrg4 and xGrg5 constructs were digested from a pGloMyc tagged vector (with BamHI and XbaI or BamHI and XhoI) and cloned into pCS2 plasmid. *Xenopus* CtBP1 coding region was sub-cloned into pCS2 from an Expressed Sequence Tag (EST) purchased from Open BioSystems (#7012549) using EcoRI.

2.3.3 Immunoprecipitation

Immunoprecipitation experiments were performed with lysates from transfected Cos7 cells. Cells were grown in a 60mm dish and transfected with a HA-tagged *Xenopus* Arx plasmid with Fugene 6 transfection reagent (Roche) as per manufacturers protocol. Cells were harvested 24hrs later and lysed using M-Per reagent (Pierce) supplemented with a Complete Protease Inhibitor Cocktail (Roche) and phosphotase inhibitors (Roche). Magenetic Protein G beads (Dynal) were used for the immunprecipitation experiments. The beads were pre-bound with mouse anti-HA antibody (Sigma) as per protocol. Pre-
bound beads (30 µl) were mixed with 150 µl of the cell lysate and 1X PBS to a final volume of 500 µl and incubated at 4°C for 1 hour on a rotator. Beads were recovered using a magnet and washed 3X with 1 ml cold 1X PBS+0.1% Tween, eluted with boiling in sample buffer, resolved on a SDS-PAGE gel, then analyzed by western blot. Endogenous CtBP was detected with a rabbit polyclonal antibody (Bethyl Labs) and endogenous Groucho was detected with a goat polyclonal TLE antibody (Santa Cruz).

2.3.4 RNA microinjection

*Xenopus laevis* embryos were injected in one blastomere at the two-cell stage with RNA encoding wildtype or mutant forms of Gal4-Arx. Gal4-Arx was co-transfected with DNA encoding the Gal4-UAS responsive promoter and a *Renilla* luciferase plasmid that served as the internal control. DNA constructs were linearized and RNA was transcribed using the SP6 mMessage High-Yield Capped RNA Transcription Kit (Ambion). Embryos were harvested at stage 10 (early gastrula stage), lysed, and cleared for luciferase assay as described in the above methods. Graphs indicate the trend of three individual experiments.

2.3.5 In situ hybridization

*Xenopus* Esg-1 [89] was isolated by PCR from an oocyte library using primers spanning the entire coding region, based on the published sequence (accession# U18775). Forward: 5’ CTCGAGGATGTCCCTCAAAACAGAC 3’
Reverse: 3’ TCTAGAGGACTTCAGTAGATAACC 5’
The PCR product was cloned into pCS2, linearized with Hind III, and transcribed with T7 polymerase to obtain a digoxygenin-labelled antisense riboprobe that was used for in situ
hybridization analysis [90, 91]. A xCtBP1 digoxygenin-labelled riboprobe was made by linearizing the pCS2/xCtBP1 plasmid with Kpn I and transcribed with T7. This riboprobe was used together with a fluorescein-labelled xArx riboprobe in double in situ hybridization analysis.

2.4 Results

2.4.1 Arx is a transcriptional repressor

Arx contains three evolutionary conserved protein domains: an octapeptide (OP), a homeodomain, and an OAR or aristalless domain [56] (Fig. 2.1A). Previous studies with other homeobox proteins [92, 93] found that the octapeptide motif can confer repression. We tested the hypothesis that Arx is a transcriptional repressor by utilizing a Gal4/UAS reporter system in which the Xenopus homolog of Arx, xArx is tethered to the Gal4 DNA binding domain and is co-transfected with a Gal4-responsive luciferase reporter construct in Cos7 cells. We found that the luciferase reporter was repressed in a dose dependent manner, demonstrating that Arx can function as a transcriptional repressor (Fig. 2.1B).

2.4.2 Arx repression is mediated in part by the octapeptide motif and the conserved L33 is important for repression

The Groucho family of co-repressors have been shown to bind and mediate repression via the octapeptide domain in proteins or protein families such as Engrailed, Goosecoid, Pax, Nkx, and Six [81]. To determine whether the repression activity of Arx required the octapeptide motif, we tested constructs containing a mutated version of the OP motif in the luciferase reporter assay. The constructs used were an N-terminal
deletion that lacked the OP and a leucine (position 33) to proline point mutation (L33P) (Fig 2.2). The L33P mutation is of particular interest because it occurs at a conserved leucine within the OP and has been found in patients with X-linked mental retardation (XLMR) [24, 54]. As expected, the OP deletion decreased the repressive activity of Arx by 2-fold (Fig. 2.3). In addition, the L33P mutation alleviated repression to a similar extent of the OP deletion, demonstrating the importance of this conserved leucine residue in repression mediated by the octapeptide (Fig. 2.3). Interestingly, the OP deletion or the L33P mutation did not alleviate repression completely, suggesting that there is an additional domain within Arx that is also required for repression.

2.4.3 Two independent repression domains are responsible for repression by Arx

Mutation or deletion of the OP did not abolish all of the repressor activity of Arx. To identify a putative repression domain, we tested a C-terminal truncation construct in the reporter assay. The truncation tested (Fig. 2.4) mimics a truncation within exon 5 that has been described in patients with XLMR and results from a frameshift at amino acid 483 [55]. In luciferase reporter studies, the exon 5 truncation construct also alleviated repression compared to Arx but not completely, suggesting that this region may also be responsible for repression (Fig. 2.5A).

To further define the region within the C-terminus responsible for repression, we made a construct in which a larger portion of the C-terminus was deleted (Fig. 2.4). This construct alleviated repression to a slightly greater extent than the exon 5 construct indicating that the region between amino acid 398 and 448 is important for repression (Fig. 2.5B). From this set of data, we conclude that Arx contains two repression
domains, the octapeptide and another domain located in the C-terminus at amino acids (398-448) of the protein, termed ORD (other repression domain).

2.4.4 The Octapeptide mediates repression via the Groucho family of co-repressors

The Groucho family of co-repressors is known to bind to transcription factors and mediate repression, specifically via octapeptide motifs [82]. Two Xenopus Groucho homologs have been reported [89, 94]. We were unable to identify additional Groucho genes by degenerate PCR experiments (data not shown). The first, xGrg4, is co-expressed with xArx in similar regions of the developing forebrain [71, 94].

The second Xenopus Groucho homolog, Esg-1 (Enhancer of Split Groucho-1) is most similar to human Tle-1 [89] however, the expression pattern has not been described. We determined that Esg-1 is co-expressed with xArx in overlapping regions of the developing Xenopus brain (Fig. 2.6). We also tested Esg-1 in the luciferase reporter assays. Because we obtained similar results in reporter gene assays for xGrg4 and xEsg-1 (data not shown), we used xGrg4 for the remainder of our experiments.

To test whether the Groucho family of co-repressors mediate repression of Arx through the OP, we co-injected wildtype and mutant Arx with xGrg4 in Xenopus embryos. When co-injected with Arx, Grg4 enhanced repression of the luciferase reporter in a dose-dependant manner (Fig. 2.7A). We also found that Grg4 did not have an effect on the L33P Arx mutant in reporter assays, again emphasizing the importance of this domain and conserved leucine in repression (Fig.2.7B). Our results agree with previous results in which the physical interaction between Arx and Grg4 was identified and shown to be weakened with the L33P mutant protein [77].
To further assess OP function, we tested an OP deletion construct (Fig. 2.2) in the reporter assay. As expected, Grg4 did not enhance repression (Fig. 2.7B). These results also suggest that repression through the ORD is not mediated through the Groucho family of proteins because neither the OP deletion nor the L33P mutation constructs, which still contain an intact ORD, showed any enhancement of repression upon addition of Grg4. Taken together, our results indicate that Arx has two independently regulated repression domains, the OP and ORD.

2.4.5 C-terminal Binding Protein 1 (xCtBP1) mediates repression of Arx

C-terminal binding proteins are co-repressors that bind the C-terminus of many transcription factors [84]. In addition, many proteins have been found to mediate repression by binding to both Groucho and CtBP [95-97]. We hypothesized that CtBP may mediate repression through the ORD in Arx. To determine whether xCtBP1 is expressed in overlapping areas of xArx expression, we performed double whole mount in situ hybridization of *Xenopus laevis* embryos (Fig. 2.8). We found that xArx and xCTBP1 are expressed in overlapping domains of the brain during later stages of development and therefore have the potential to interact molecularly. We next tested whether repression by Arx could be enhanced by the co-transfection of CtBP1 in a reporter assay. Similar to the results for Grg4, CtBP1 enhances repression of the reporter ~3-fold compared to Arx alone (Fig. 2.9A). To determine whether xCtBP1 can physically interact with Arx, we performed co-immunoprecipitation experiments with Cos-7 cells transfected with HA-tagged Arx (Fig. 2.9B). HA-Arx precipitated endogenous CtBP, suggesting that CtBP1 mediates Arx repression via physical interaction.
2.4.6 The ORD mediates repression by CtBP1

To determine if the ORD was responsible for mediating repression via CtBP, we tested the exon 5 and the 398 truncation constructs in a co-transfection reporter assay with CtBP. We found that CtBP was able to enhance repression of these constructs (Fig. 2.10A).

Because we observed enhanced repression with the C-terminal truncation constructs, we next tested constructs that had a C-terminal truncation together with the L33P mutation. Upon co-transfection with the mutated Arx, CtBP1 could still enhance repression (Fig. 2.10B). Previous studies report that Arx containing the L33P mutation can weakly bind Groucho [77]. We postulated that the observed repression could be due to Grg4 interaction with the mutated OP and that CtBP may be able to work through this complex to enhance repression. To address this possibility we made constructs with the truncated C-terminus and the OP deletion (Fig. 2.11A). CtBP1 could enhance repression of the -OP/exon5 truncation construct but not of the –OP/398 truncation construct, validating that CtBP1 can mediate repression through the ORD (Fig. 2.11B). This data also implicates an interaction or complex formation between Arx, Groucho, and CtBP.

The Arx protein contains a PLGLS motif within the 398-448 amino acid region of the ORD (Fig. 2.12A) that conforms loosely to the defined CtBP binding site [84]. We tested a construct that deleted this motif and the OP in the luciferase reporter assay. We found that CtBP could still enhance repression of Arx (Fig. 2.12B). Although the involvement of the PLGLS motif cannot be ruled out, the data suggests that sequence outside of this motif is necessary for CtBP to mediate repression through the ORD.
2.4.7 *Arx interacts with both Groucho and CtBP*

To investigate the possibility of an interaction between Arx, Grg4, and CtBP, co-immunoprecipitation experiments were performed using Cos 7 cells that were untransfected or transfected with HA-Arx (Fig. 2.13). The results indicate that endogenous CtBP could precipitate with endogenous Grg4 only in the presence of Arx. This result also indicates that Arx can interact with both co-repressors in a complex, however it is not clear whether the two co-repressors interact directly or indirectly. Collectively, this data suggests that Arx can physically interact with both Groucho and CtBP since both co-factors could be precipitated in the presence of Arx. This result further supports the evidence that Arx is a transcriptional repressor regulated by both Groucho and CtBP.

2.4.8 *Arx contains additional conserved domains.*

Arx contains regions of conservation outside of the homeodomain and OP (see Chapter 1). To determine whether these other conserved regions had an effect on Arx function, the constructs were tested in a luciferase reporter assay. Mutations were made that deleted conserved regions 1, 2, or 3, or the OAR domain (Fig. 2.14A). It was found that none of these mutations affected the function of Arx as compared to the control (Fig. 2.14B).

2.5 *Discussion*

Arx plays a critical role in brain development and mutations that result in an altered protein cause neurological disorders. Despite the wealth of information known about the clinical manifestations of mutations in Arx, little is known about the molecular
consequences of these mutations. In addition, it is not completely understood how Arx acts to repress its downstream targets. This study suggests that repression by Arx is complex and requires more than one co-factor.

In this work, it was found that the octapeptide motif is important for repression by the co-repressor Groucho. Also, the L33P mutation that is found in patients with XLMR [24, 54] disrupts Groucho mediated repression via the OP. Our results agree with previous findings in which it was shown that the L33P mutation in the Arx OP weakens the physical interaction with Groucho [77].

Through reporter assays with the exon 5 truncation construct, which in humans results in X-linked mental retardation with abnormal genitalia (XLAG) [55], it was determined that Arx contains a second repression domain (ORD). The ORD was further narrowed down to a region in the C-terminus corresponding to amino acids 398-448. Furthermore, the reporter assay data confirms that CtBP mediates repression of Arx though the ORD.

In some of the assays that were performed in embryos, variable were obtained when compared to those results from experiments done in transfected Cos7 cells (Fig. 2.15). Although the values of the reporter assay done in embryos differed from experiment to experiment, which could be due to genetic differences in each batch of fertilized embryos, a consistent trend of activity was observed from several individual experiments.

The differences that were obtained from cells compared to embryos could be due to the in vivo nature of the embryos as compared to cultured cells. Embryos are made up
of thousands of cells that are not identical, whereas Cos7 cells are one cell type. In addition, Cos7 cells are kidney cells and have a different cellular environment and may contain factors that act to alter the function of Arx as compared to the factors present in the embryo during early development. Future studies in different cell types would provide insight into this possibility. In any case, this work provides evidence for the presence of two independent repression domains in Arx that are regulated by the co-repressors Groucho and CtBP.

The co-repressors Groucho and CtBP are shown to mediate repression by qualitatively different mechanisms; Groucho is involved in long-range repression by influencing enhancers that are located thousands of base pairs away and CtBP in short-range repression by interfering with activators at a nearby locus [78]. Recently, several transcriptional repressors have been reported to interact cooperatively with both Groucho and CtBP. It has been shown that the *Drosophila* repressor proteins Brinker, Hairy, and Single-minded can mediate repression of downstream genes by interacting with Groucho, CtBP, or both [95-97]. In addition, the *Drosophila* Hairless protein has been shown to confer repression of the Hairless-Su(H) complex via binding both CtBP and Groucho in combination [98].

In this work, we demonstrate that Arx can physically interact with both Groucho and CtBP because both co-factors were precipitated in the presence of Arx. Transcriptional repressors such as Brinker, Hairy, Hairless, and Single-minded, repress downstream targets in a context-dependant manner in which certain promoters require different combinations of CtBP and Groucho [95-98]. This type of regulation could also
be a mechanism for repression by Arx, however additional studies with downstream targets are necessary to test this possibility.

Previous studies of Arx deficient mice have shown that Arx is important for many developmental processes. It has been shown that expression of Arx is crucial for the proper migration of GABAergic interneurons in the mouse cortex and has an important role in neuronal cell differentiation and proliferation [53, 66]. Additional loss of function studies indicate that in the mouse pancreas, Arx is important for proper endocrine cell specification through the inhibition of Pax4, in which Arx promotes the α-cell fate and antagonizes β- and δ-cell commitment [57]. Recently, it has been described that a transcription factor involved in cell differentiation and cell fate commitment, Tcf3, mediates its repressive effects in these processes by partnering with Groucho and CtBP [99]. It is interesting to speculate that the role of Arx in proliferation, differentiation, and cell fate specification could be regulated by the interaction with Groucho and CtBP in a similar manner as Tcf3.

Also of interest are the findings in Drosophila that show that Groucho and CtBP are expressed in non-overlapping regions in early development and that CtBP may act temporally to impose tight regulation or fine tuning of repression at certain key developmental stages [97]. In our report we show that the expression of Arx and CtBP in Xenopus embryos does not overlap until later stages of development (Fig. 2.8). This finding indicates that there could be a temporal effect of CtBP mediated repression of Arx, however, further experiments are necessary to look into this possibility.
Our results indicate that the conserved regions in Arx including the OAR do not affect its function as a transcriptional repressor (Fig. 2.14). It is possible that these regions do play a role in Arx function, however in our assay we did not observe it. Further work in dissecting the Arx protein is necessary to determine the potential role of these domains in the function of the protein.

In this work we set out to determine how disease-associated mutations affect the molecular function of the transcriptional repressor, Arx and we show that two disease-associated mutations affect Arx repression. We provide evidence that Arx functions as a strong repressor in which both Groucho and CtBP co-repressors modulate repression through two distinct domains in the protein. This data supports the recent reports that demonstrate that repression is an important event during many developmental processes and in some instances several different co-repressors are necessary to regulate transcriptional repression. Future work aimed at determining the downstream targets of Arx in the brain will provide a better understanding of the mechanisms involved in their regulation and how mutations that result in perturbation of repression result in disease.
Figure 2.1. *Xenopus* Arx functions as a transcriptional repressor. A. Schematic representation of Arx conserved domains. B. Arx functions as a transcriptional repressor. Luciferase assay performed using lysates from Cos 7 cells co-transfected with a Gal4 DBD-Arx expression plasmid and a Gal4-UAS-Luc reporter plasmid. All values are significant to the control.*p<0.05

**OP**: Octapeptide motif, **HD**: Homeodomain, **OAR**: Orthopedia-Aristaless-Rx Domain **CR**: Conserved Region
Figure 2.2 Arx mutation constructs. A. Schematic representation of constructs used in luciferase assays.
A. Disruption of the OP results in reduced repression activity. Luciferase assay performed using lysates from Cos 7 cells co-transfected with Gal4 DBD-Arx-OP or L33P expression plasmids and a Gal4-UAS-Luc reporter plasmid. The L33P values are significantly different than the wildtype values.*p<0.05  

B. Luciferase assay performed using lysates from stage 10 Xenopus embryos micro-injected with 10pg of RNA encoding Gal4-DBD-Arx-OP or L33P and a Gal4-UAS-Luc reporter plasmid. The graph represents a trend observed from three experiments.

Figure 2.3. The octapeptide domain is necessary for repression.
Figure 2.4. Arx mutation constructs. A. Schematic representation of mutation constructs used in luciferase assays.
Figure 2.5. Two repression domains are responsible for repression by Arx. A,B. A second repression domain is located in the C-terminus of Arx. Luciferase assay performed with lysates from Cos7 cells co-transfected with 10 ng of Gal4 DBD-Arx L33P, an Exon5 truncation, or an amino acid 398 truncation expression plasmid and a Gal4-UAS-Luc reporter plasmid. Values are significant as compared to wildtype. *p<0.05
2.6. Whole Mount in situ hybridization of xEsg-1. A-D. Whole mount in situ hybridization showing expression of *Xenopus* Groucho homolog Esg-1 in the brain of varying stages of *Xenopus* embryos at stages 21, 27, 32, and 35. Ey= eye, fb= forebrain, mb= midbrain. Expression is observed in the forebrain, midbrain, eye, and somites of the developing embryo.
Figure 2.7. The Octapeptide mediates repression via the Groucho family of co-repressors. A. The Xenopus Groucho homolog Grg4 enhances the repression of Arx. Luciferase assay performed using lysates from stage 10 Xenopus embryos that were co-injected with RNA encoding Gal4-DBD-Arx and xGrg4 and a Gal4-UAS-Luc reporter plasmid. The graph represents the trend observed from three experiments. B. Luciferase assay performed using lysates from Cos7 cells co-transfected with Gal4 DBD-Arx, Gal4 DBD-Arx L33P, or Gal4 DBD-Arx-OP and xGrg4 expression plasmids along with a Gal4-UAS-Luc reporter plasmid. Values are significant between Arx with and without Grg4. *p<0.05
Figure 2.8. Double Whole mount in situ hybridization of xArx and xCtBP1. A-C. Double whole mount in situ hybridization showing overlapping expression of xArx and xCtBP in the developing brain of Xenopus embryos at stages 21, 32, and 35. Ey=eye and fb= forebrain. Arx is stained magenta, xCtBP is stained turquoise, and areas of overlap are purple. xArx expression is primarily observed in the telencephalon of the developing embryo whereas xCtBP expression expands from the eye to the brain as the embryo matures.
Figure 2.9. C-terminal Binding Protein 1 (xCtBP1) mediates repression of Arx. A. Luciferase assay performed using lysates from Cos7 cells co-transfected with Gal4 DBD-Arx and xCtBP expression plasmids and a Gal4-UAS-Luc reporter plasmid. *p<0.05 B. Co-immunoprecipitation of CtBP performed using lysates prepared from Cos7 cells transfected with HA-Arx. Control IP was performed using mouse IgG.
Figure 2.10. CtBP enhances repression in constructs with an intact OP. A.B. Luciferase assay performed using lysates from Cos7 cells co-transfected with Gal4 DBD-Arx, Gal4 DBD Arx L33P, Gal4 DBD-Arx-Ex5, or Gal4 DBD-Arx 398 truncation with or without the addition of CtBP. B. Constructs have a mutation at L33P together with a C-terminal truncation. All values are significantly different from each other upon the addition of CtBP. *p<0.05.
Figure 2.11. The ORD mediates repression by CtBP1. A. Schematic representation of mutation constructs used in luciferase assays. B. Luciferase assay performed using lysates prepared from Cos7 cells co-transfected with Gal4 DBD-Arx, Gal4 DBD-Arx-OP/Ex5 truncation, or Gal4 DBD-Arx-OP/398 truncation and xCtBP expression plasmids along with a Gal4-UAS-Luc reporter plasmid. Values are only significant between wildtype or Ex5 constructs with and without CtBP. *p<0.05.
Figure 2.12. Deletion of the PLGLS motif does not disrupt repression by CtBP.  A. Conservation of Arx protein sequence in human, mouse, Xenopus and Zebrafish indicating the region containing the ORD and the putative CtBP binding site within it. B. Luciferase assay performed using lysates from Cos7 cells transfected with Gal4 DBD-Arx, Gal4 DBD-Arx-PLGLS with or without the addition of CtBP. Values are significant to each other upon the addition of CtBP. *p<0.05.
Figure 2.13. Arx interacts with Groucho and CtBP. A. Co-immunoprecipitation of Groucho performed using lysates prepared from Cos7 cells transfected with HA-Arx. Control IP was performed using mouse IgG.
Figure 2.14. Arx contains additional conserved regions. A. Schematic representation of Arx conserved domains. B. Arx functions as a transcriptional repressor. **OP:** Octapeptide motif, **HD:** Homeodomain, **OAR:** Orthopedia-Aristaless-Rx Domain **CR:** Conserved Region B. Luciferase assay performed using lysates from stage 10 *Xenopus* embryos micro-injected with 10pg of RNA encoding Gal4 DBD Arx-OAR or CR1-3 and a Gal4-UAS-Luc reporter plasmid. The graph represents a trend observed from three experiments.
Figure 2.15. Arx contains two independent repression domains. Luciferase assay performed using lysates from stage 10 Xenopus embryos micro-injected with 10pg of RNA encoding Gal4 DBD-Arx L33P, Arx-Ex5, or the double mutation with a Gal4-UAS-Luc reporter plasmid. The graph represents a trend observed from three experiments.
CHAPTER 3

ARX BINDING SITE IDENTIFICATION

3.1 Abstract

Homeobox genes encode transcription factors that are highly conserved from *Drosophila* to humans and are crucial for many developmental processes. The homeodomain (HD) has been studied extensively and found to have a helix-turn-helix structure that has the ability of binding to preferential DNA sequences. The DNA sequences to which the homeodomain can bind depend on key residues within the homeodomain itself. These residues also dictate whether the HD binds to DNA target sites as a monomer, a dimer, or cooperatively with other homeodomain proteins. The *Aristaless* related homeobox protein, Arx, is a member of the paired-type homeodomain family of proteins and is important for brain development. Mutations in Arx cause severe brain malformations. Many of these mutations are found to occur within the homeodomain. It is not known how these mutations affect the function of Arx as a transcription factor because a neural downstream binding site has yet to be determined. In this report we describe our efforts to identify a binding site for Arx. Our results indicate that Arx does not bind to the sites previously described for other paired-type
homeodomain proteins and that further experiments are necessary to identify a binding site.

3.2 Introduction

Homeodomain proteins are transcription factors that play a fundamental role in many aspects of development. The homeodomain was first discovered in *Drosophila* as a conserved element within homeotic selector genes and was identified as a 180 base pair region that encodes a helix-turn-helix DNA binding motif [2, 100]. Since then the molecular evolution and phylogenetic analysis of the homeodomain has been reported. This analysis allowed for the classification of the homeodomain proteins based on functional and structural characteristics [101]. Additionally, many studies have been done to determine the preferential DNA binding sites of the different classes of homeodomain proteins [13]. Recently, it has been reported that homeobox genes expressed in the forebrain control many aspects of brain development in both mouse and humans. It was found that mutations often occur in these genes and while the molecular functions of the genes are known in many cases, the downstream targets are not [25].

The *Aristaless* related homeobox gene, Arx is an important transcription factor that plays a critical role in forebrain development. At least fifty-nine mutations in Arx have been identified to date that cause various forms of mental retardation [24]. Many of these disease-causing mutations occur within key residues of the homeodomain, suggesting that they disrupt the DNA binding function of this domain. Elucidation of an Arx downstream target would help to understand how mutations within the homeodomain result in disease.
Arx belongs to the paired-type homeodomain group of proteins and contains a Q50 residue within its homeodomain. Paired-type homeodomains have been shown to bind as dimers cooperatively to sites consisting of two palindromic TAAT sequences that are separated by two or three random nucleotides, P2 or P3 sites [20]. It has been further shown that Q50 homeodomains prefer P3 sites over P2 sites [13]. In addition to binding as homodimers, it is known that members of the paired-type homeodomain family can also bind to their preferential DNA binding sites as heterodimers with other homeodomain proteins of the same or different class [102]. Despite what is known about Arx and the homeodomain itself, an Arx binding site has yet to be reported. In this study, we employed two different methods to identify a DNA binding site that could be used in downstream reporter assays to address the consequence of disease-associated mutations in the homeodomain on Arx function as a transcription factor.

3.3 Experimental Procedures

3.3.1 Electromobility Shift Assay (EMSA)

The Xenopus Arx gene [71] was cloned into Gal4 DBD/PCS2 vector (as described in Chapter 2) and translated in vitro (IVT) with the TNT SP6 wheat germ lysate Quick coupled transcription/translation kit (Promega). The Luciferase control plasmid that is included with this kit was also translated and served as the negative control protein for these assays.

Oligonucleotides (Fig. 3.1) were annealed and their 5’ ends labeled with $[^{32}P]ATP$ and T$_4$ polynucleotide kinase (Invitrogen). The binding reaction (based off of Kimura et al, 2000) was carried out in a 20µl reaction containing 12mM HEPES (pH
7.9), 12mM KCl, 0.6mM MgCl₂, 1.2mM dithiothreitol, 10%glycerol, 2µg of poly (dI-dC), 5 fmol of the ³²P-labeled probe, and 2 µl of IVT protein. In the competition assays, 100-fold molar excess of unlabelled competitor DNA was added to the standard mixture. After incubation on ice for 30 minutes, the samples were loaded onto a 5% polyacrylamide gel in 0.5X Tris borate-EDTA (TBE) buffer, and electrophoresed at 150V/cm for 2-3 hours at 4°. The gel was then dried and autoradiographed at -80° overnight.

For the EMSA with the Al/C11 probe, the procedure was the same as above but with binding conditions to that of [103]. Also, the EMSA using the Pax4 probe was done according to [70]. The primers used to amplify the 200bp region in the Pax4 promoter were:
F: GAGTCAGACAGTGAAGGCAG
R: CCCATAACCTACCCCTCTCTC

Modifications that were made to EMSA experiments included trying different non-specific competitors to remove the binding observed in the negative control sample. Different reagents used were salmon sperm and dAdT in place of dIdC.

3.3.2 Systematic Evolution of ligands by exponential enrichment (SELEX)

SELEX experiments were carried out for the most part as described by Kojima et al. Primers used were as follows:
Selex N18:
GATTTTGCGAATT CATGCTGCA-(N18)- CTCTGACAAGCTTTGCCATTGG
Selex R: CCAATGGCAAAGCTTTGTCAGAG
Selex F: GATTTTGCGAATTCCATGCTGCA

Binding Buffer: 10mM Tris-HCl (pH7.5), 50mM NaCl, 1mM EDTA

A random pool of oligos was made from synthesizing the N18 oligonucleotide. The random oligos were made double-stranded by PCR reaction using 250 pmol of Selex N18 and SelexR primers (94°C for 3 minutes, 55°C for 5 minutes, 72°C for 10 minutes) and then purified using a nucleotide removal kit (Qiagen). In the first round of the selection procedure was started by adding 5pmol of random oligos to 5µl of in vitro-translated HA-Arx protein in Binding Buffer containing 0.2 μg /µl BSA, 50ng/µl dl-dC, and 0.05% NP-40 to a final volume of 30µl. This mixture was incubated for 30 minutes at room temperature. In vitro translated HA-Arx that lacked the homeodomain (which was removed using Sma I sites) served as the negative control.

After incubation, 12µl of Protein G magnetic beads (Dynal) that were pre-bound with anti-mouse-anti-Ha antibody were added to the mixture and agitated gently for 20 minutes at room temperature in a total volume of 100µl by the addition of Binding Buffer. The tubes were then placed in a magnetic holder and washed three times with 1ml of Binding Buffer. The DNA was then eluted by heating at 95° for 5 minutes and the supernatant kept for amplification.

DNA was amplified using 5µl of the eluate and 60pmol of the Selex F and Selex R primers using standard PCR parameters for a number of cycles that produced a band in the experimental lane and not in the negative control lane. An additional 120 pmol of the Selex primers were added and the mixture was subjected to a further round of PCR (95°C
for 3 minutes, 55°C for 5 minutes, and 72°C for 10 minutes. The PCR product was purified with the Nucleotide Removal Kit (Qiagen). This product was then subjected to four more rounds of amplification for a total of five rounds. Then the product was cloned into the pCR2.1 TOPO vector (Invitrogen) and the sequence analyzed.

Modifications to wash buffers were made in an attempt to remove non-specific binding. The modified buffers that were tried were: 1XPBS with 0.01% Tween, 1XPBS with 0.01% Tween and 300mM NaCl or 500mM NaCl. Another modification that was tried was a random oligo library that contained (AAT) in the middle of the random sequence:

\[
\text{GAATT TTGCGAATT CATGCTGCA-N7-AAT-N7-CTCTGACAAGC TTTGCCATTGG.}
\]

In addition a His-tagged Arx homeodomain was purified from bacteria and used in the assay. The xArx homeodomain was isolated from the Gal4-Arx plasmid (Chapter 2) and cloned in frame with the His tag into the pRSET B vector (Invitrogen) and purified using B-PER 6X-His fusion protein purification kit (Pierce).

3.4 Results

3.4.1 Arx does not bind a typical paired-type homeodomain DNA binding motif

Homeobox proteins are known to be key regulators in neural development, however most of the downstream genes involved in these developmental pathways remain to be identified. Given the importance of Arx in brain development, we set out to identify an Arx DNA binding site that could be used in further applications to study the
effects of disease-associated mutations that occur within the homeodomain on Arx function.

To identify an Arx binding site, we used sites that were published for other members of the paired-type *Aristaless* related homeobox protein family. Protein-DNA interactions were examined by utilizing the Electromobility Shift Assay (EMSA) using double-stranded oligonucleotides for each of the potential binding sites. The oligonucleotides were labeled and used as probes that were added to lysate containing in vitro translated *Xenopus* Arx protein. The first site tested was the photoreceptor conserved element-1 (PCE-1) site. The PCE-1 site is found in genes such as the photoreceptor arrestin promoter in which its activation is stimulated by the retinal homeobox protein (Rx) in an *in vitro* reporter assay [104]. Arx is a close family member of the Rx protein. We found that Arx did not bind to the PCE-1 site specifically as compared to the positive control protein sample, Rx (data not shown).

Binding sites for other Arx family members were also tested by EMSA. Among these were binding sites for: Prx2, Cart I, and Pax4 [70, 103, 105]. Prx2 and CartI are both paired-type homeodomain-containing proteins that belong to the *Aristaless*-like family of proteins and are important for craniofacial and limb development [13]. Prx2 has been shown to bind a TAACTAATTAAC site in reporter and EMSA assays [105, 106]. CartI has been reported to bind a consensus sequence that differs from the palindromic P3 sequences previously reported [107]. It has been shown by EMSA that CartI can bind to a (T/C)TAATTAA(T/A)(T/A) sequence. CartI can bind this site to an even stronger extent when in a complex with a vertebrate homolog of *Drosophila* CII, Hox11L1 or Tlx2.
protein [103]. It was also shown that the *Drosophila* Aristaless (*al*) could also bind this sequence when in a complex with Hox11L1 [103]. We tried to confirm this finding with *Xenopus* Arx but were unsuccessful (Fig.3.2). However, CartI was used as a positive control and could bind to the site confirming that our assay was working. Our EMSA results also showed that Arx was unable to specifically bind to the Prx2 site (data not shown). From this data, we conclude that Arx does not bind to the same DNA binding sites as other proteins in its family.

It has recently been reported that Arx and Pax4 are able to act as transcriptional repressors to inhibit expression of one another to mediate proper endocrine cell fate in the pancreas [70]. It was shown by EMSA that Arx could bind a site within the Pax4 promoter. This finding was further confirmed by ChIP analysis and luciferase reporter assays [70]. We located the nucleotide sequence in the region that was published for this promoter [70] and tried this site in our EMSA experiments. However, this site also produced negative results in our experiment.

Arx is a member of the paired-type homeodomain protein family and contains a Q50 residue in its homeodomain. Q50 homeodomains generally prefer P3 sites of the sequence TAAT-YNR-ATTA, where Y is either C or T and R is either A or G [13]. In addition, a (TAAT-NNN-ATTA) palindromic site (Table 3.1) where N is a random nucleotide was also tested by EMSA (Fig.3.3). Our results indicate that Arx could not bind to this site specifically.
3.4.2 *SELEX* binding site identification used as an alternative approach to locate an Arx binding site

In a further effort to identify an Arx binding site we employed the use of the systematic evolution of ligands by exponential enrichment (SELEX) experiment as an unbiased approach to identifying an Arx binding site that could be used in downstream applications. In vitro translated HA-tagged Arx protein was pre-bound to magnetic protein G magnetic beads and mixed with a solution of random double-stranded oligonucleotides. The oligos contained a flanking sequence on either end that served as primer binding sites for PCR amplification. As a negative control, HA-tagged Arx that lacked the homeodomain was used. After 5 rounds of enrichment and washes, the DNA product was eluted from the beads and amplified. The resulting PCR products were cloned and sequenced. The analyzed sequence revealed a possible consensus sequence. The consensus sequence, as well as the six sequences from which it was derived, were used in the EMSA experiments (Fig.3.4). However, none of the possible sequences showed specific binding to Arx in the EMSA.

The possible consensus sequences did not bind to Arx specifically. One reason for this may be that oligos were binding to the beads in a non-specific manner. We tried several different approaches to combat this problem. First, we used a more stringent, higher salt wash buffer to effectively wash away any non-specific binding. To focus our strategy, we also tried a new set of random oligos that contained the sequence (AAT) in the middle of the random oligo. We did this to select for sequences that would contain a partial sequence of the binding motif that has been shown to bind paired-type
homeodomain proteins [13]. This would allow for flanking regions outside of the AAT that are important for Arx binding to be identified. Despite these efforts, we did not observe specific binding.

Due to the fact that the previous attempts did not yield a specific binding site, we decided to use a purified protein in the SELEX experiments. We made a HIS-tagged Arx construct that contained only the homeodomain since we were not able to produce full-length Arx in the bacteria. This purified protein was used as the starting material for the SELEX experiments. As a negative control we used an N-terminal portion of Arx that lacked the HD. After just one round of SELEX we saw a band amplified for our negative control indicating that a specific binding site for Arx was not selected.

3.5 Discussion

Homeobox genes have been long determined to be important regulators of development. Homeodomain proteins typically serve as transcription factors that regulate downstream genes by the ability of the homeodomain to recognize sequence-specific DNA binding motifs of downstream target genes in a spatial and temporal manner [1]. We set out to identify a DNA binding site for the *Aristaless* related protein, Arx. Arx is important for normal brain development and many human mutations have been reported to occur within the homeodomain that cause severe brain malformations. These mutations are speculated to hinder homeodomain function, however a downstream target has not yet been identified to test this hypothesis.

Arx is a member of the paired-type family of homeodomain proteins. It contains a Q50 residue within the homeodomain that typically indicates binding to a particular
TAAT containing target site [11]. It has also been reported that members of the paired
class of homeodomain proteins are able to dimerize and bind to a TAAT-x-ATTA
palindromic (P2 or P3) sequence where x is a variable number of residues that depends
on the amino acid present at position 50 of the homeodomain [20]. Members of the
paired class that contain a Q50 in their homeodomain are known to preferentially bind to
P3 sites [20]. Other members of the paired family have been shown to bind these types
of sites [13].

We tested the hypothesis that Arx should also bind a similar sequence by testing
previously described sequences for *Aristaless* family members by Electromobility Shift
Assay (EMSA). Sites tested included those known to bind the retinal homeobox (Rx)
protein, Prx2, and Cartl as well as a P3 site. After several attempts and modifications to
the protocol, our results indicate that Arx was not able to bind to any of the sites in our
assay. Binding to a P3 palindromic site is usually done by a protein that can homo or
heterodimerize [20]. We do not have any data suggesting that Arx can homo- or
heterodimerize, however it remains a possibility and should be addressed in future
experiments.

It is quite unexpected that Arx was not able to bind any of the tested sites given
the fact that the Arx homeodomain contains conserved residues known to be important
for binding TAAT containing sequences: Arginine (R position 5), isoleucine or valine
(I/V position 47), asparagine (N position 51), and a basic residue at positions 2 or 3 [11]
(Fig.3.5). One possibility is that there was another protein present in the lysate that could
bind the sites we were testing. This makes some sense in that non-specific binding was
seen in every control lane in our EMSA experiments indicating that there may have been a protein in the in vitro translated lysate that was binding to the site and masking the specific binding of Arx (Fig.3.2). To circumvent this problem, a purified protein could be used instead of in vitro translated protein, which contains other factors that are present in the lysate.

Another possibility is that Arx could need a partner in order to efficiently bind to its target sites and that partner was not present in our assay. It has been reported that members of the paired class of homeodomain proteins bind to palindromic TAAT sites in cooperation with other homeodomain containing proteins [20]. In fact, this is thought to be the reason why homeodomain proteins that bind to similar target sequences are able to exert different molecular functions. However, if a partner protein were necessary for Arx binding to its target site, it may not be present in purified protein lysate. To ensure that the proper partners were available, a neural specific lysate from *Xenopus* brains at varying stages of development could be used in an EMSA. An antibody for Arx could be used to super-shift the complex that contained Arx to show that the binding was specific. Further experiments would then be necessary to identify the potential cooperative partners. In addition, it has been speculated that protein-protein interactions that involve the homeodomain could allosterically hinder the specificity of binding by the helix-turn-helix region of the homeodomain, and thereby change the degree of specificity or the preferred site to which it binds [108]. This situation suggests how a protein that partnered with Arx could affect its DNA binding.
Yet another possibility is that Arx does not bind the typical paired-type DNA binding sequences that have been reported. It has been shown that paired-type homeodomain proteins can dimerize on P2 sites and that residue 9 of the recognition helix can alter the cooperative binding of the homeodomains [20]. The residues that have been studied at this position are serine, glutamine, or lysine [20]. Arx contains a threonine at this position. Although threonine is similar to serine because it also contains an alcohol group, it is possible that this alteration could result in a difference in the DNA binding preference of Arx. It could also indicate that Arx may not dimerize on the sites we tested and therefore a half site could be tested to address this possibility. It could also be that Arx binds to a different site that has not been described.

Since our results indicated that Arx might bind a different site than its related family members, we employed the use of a random binding site assay (SELEX) as an unbiased approach to identify a binding site. After several attempts and modifications to the protocol, we still were not able to detect any specific binding. Our EMSA experiments indicate that Arx may need a partner to efficiently bind to DNA, therefore, it is possible that the SELEX experiments did not work because a partner protein was not present.

An alternative to this approach to SELEX would be to make use of microarray and chromatin immunoprecipitation (ChIP) technology. It would be interesting to compare RNA recovered from embryo lysates in which Arx protein expression was overexpressed or down-regulated to identify potential downstream targets. A neuronal specific array could have been used in order to look for just those genes expressed in the
brain and important for cognitive function. A model for the genetic network in the pancreas of which Arx is involved has been reported [70]. This information could be used to help narrow down the possible targets generated from the microarray data since most of those genes are also expressed in the brain. Additionally, chromatin immunoprecipitation (ChIP) of Arx from a neuronal lysate could be used in which the precipitated product could be further analyzed by microarray (ChIP on Chip) to identify the target genes. These approaches would not only aid in identifying a binding site, but also downstream target genes of Arx.

Previous results from our lab have implicated Dlx5 as a possible downstream target of Arx. Overexpression of a constitutive activator form of Arx, xArx-VP16, in Xenopus embryos revealed ectopic expression of Dlx5 [76]. Dlx genes are known to be involved in regulating the differentiation of GABAergic neurons of the telencephalon [53, 109]. Arx has been shown to be involved in the migration of GABAergic neurons [53]. It has been speculated that Arx and Dlx genes are involved in the same genetic pathway [108]. We used a region of the Dlx5 promoter in a reporter assay with Arx, however no activity was observed. Recently, it has been shown that Arx can bind the Pax4 promoter in the pancreas [70]. We also tried this site in our EMSA experiments and were unable to detect any specific binding. It has also been recently reported that Arx can synergize with MyoD and Mef2C to activate the muscle-specific transcription of the myogenin promoter, although it is not certain that Arx binds the promoter directly [110]. Collectively, these reports together with our findings that Arx may bind a non-typical site
in addition to alternative approaches will help to identify an Arx target gene and possibly a mechanism for binding by the Arx homeodomain.

Identifying a target promoter sequence will lead to a better understanding of how Arx exerts its transcriptional functions on downstream genes involved in brain development and how aberrations in the Arx homeodomain contribute to diseases such as XLMR. Very recently three direct neural downstream targets of Arx were identified. Felp et al used a genome-wide mircoarray analysis of subpallia micro-dissected from E14.5 Arx deficient mice to identify possible downstream targets [111]. Overall, 84 genes were found to be differentially regulated in the mutant mice as compared to wildtype. Of these Ebf3, Lmo1, and Shox2 were validated by EMSA, ChIP, and reporter assays as being direct downstream targets of Arx.

Fulp et al also determined the consensus binding site for Arx as: TAATTA. This site contains a TAAT motif, which has been shown to be a sequence that paired-type homeodomain proteins preferentially bind and further demonstrates the importance of the conserved residues of the recognition helix. Two of the sites tested in this dissertation contained the TAATTA sequence, however the flanking regions on either side of the site are different than those reported by Fulp et al. It may be that the flanking residues are important for binding and could be why binding was not observed in the EMSA experiments of Chapter 2. In addition to utilizing different techniques, Fulp et al also used a purified protein representing only the homeodomain of Arx in their EMSA experiments. Despite the fact that a purified homeodomain-only Arx protein was used in the SELEX experiments, it could be another factor that resulted in the lack of specific
binding in the EMSA experiments of Chapter 3 in which an *in vitro* translated protein lysate was used.

The binding site that was identified by Fulp *et al* is not palindromic like a P3 site, which suggests that Arx can bind DNA as a monomer, and could be why Arx did not bind to any of the palindromic sites that were tested in Chapter 3. This finding correlates with the reasoning mentioned earlier in this chapter indicating that Arx contains a non-conserved residue at position 9 in the recognition helix that may alter its dimerization abilities. Although, it seems Arx can bind DNA as a monomer, there is still a possibility that Arx can heterodimerize with other homeodomain proteins to bind additional sites that have not been described. Analysis of the other transcription factor binding sites (near the Arx site) within the identified targets would indicate what other factors can bind the target promoter and might provide clues into potential Arx partner proteins. Future information about the binding preferences of Arx together with the discovery of more targets will provide insight into this possibility.
Table: Oligonucleotide Probes Used in the EMSA Experiments

<table>
<thead>
<tr>
<th>PROBE</th>
<th>SEQUENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCE-1</td>
<td>GATCCAAGCTTTCAATTAGC</td>
</tr>
<tr>
<td>P3</td>
<td>CCTGAGCCTAAATCCGATTCGTGAC</td>
</tr>
<tr>
<td>Prx2</td>
<td>GCTAACTAATTAACGC</td>
</tr>
<tr>
<td>Al/CII</td>
<td>AAAGTTTCTTTAATTAATGCTAAG</td>
</tr>
</tbody>
</table>

**Figure 3.1. EMSA Probes.** A. Oligonucleotide probes used in the EMSA experiments. Bold letters indicate the typical P3 palindromic site.
Figure 3.2. Al/Cll EMSA. A. EMSA with Al/Cll oligonucleotide probe. Assay was conducted using 2µl IVT protein. Lanes 1-3 are control protein. Lanes 4-6 are Cart1 protein. Lanes 7-9 are Arx protein. No competitor (-), un-labeled non-specific competitor (ns), or un-labeled specific competitor (s) were added at a 100-fold molar excess to each set of reactions. Arrow indicated specific binding.
Figure 3.3. P3 EMSA. A. EMSA with P3 oligonucleotide probe. Assay was conducted with 2µl IVT protein. Lanes 1-3 are control protein. Lanes 4-6 are Arx protein. No competitor (-), un-labeled non-specific competitor (ns), un-labeled specific competitor (s) were added at a 100-fold molar excess to each set of reactions. Arrow indicates specific binding.
Figure 3.4. Consensus Sequences obtained from SELEX. A. The seven oligonucleotide probes out of a possible twenty that were used in EMSA experiments. B,C. Multiple expectation maximization for motif elicitation (MEME) illustrating the analysis used to obtain the consensus sequences from the SELEX assay.
Figure 3.5. **Homeodomain schematic** (adapted from Wilson *et al.*, 1996). A. Arx homeodomain sequence and helix locations. B. Schematic of the homeodomain residues that contact the TAAT core motif. The circled w indicates a water molecule.
CHAPTER 4
RX N-TERMINAL FUNCTIONAL DISSECTION

4.1 Abstract

The retinal homeobox gene (Rx) is member of the paired-type homeodomain protein family and contains a conserved octapeptide motif. Rx plays a critical role in vertebrate eye development however, despite what is known about Rx in eye development, there is little reported about its molecular function. Previous results have shown that Rx is a weak activator of transcription, however Rx contains an octapeptide motif, which is known to confer repression. In this report, we found that Rx acts as a weak activator of the *Xenopus* rhodopsin promoter and has increased activity when the octapeptide is deleted. Furthermore, we found that a portion N-terminal to the octapeptide contributes to the activity of Rx. This data illustrates that the N-terminal portion of Rx may be involved in the inhibition of octapeptide function.
4.2 Introduction

The retinal homeobox gene (Rx) encodes a paired-type homeodomain protein and is a member of the Aristaless-related family of homeodomain proteins. The Rx protein contains a highly conserved homeodomain, octapeptide (OP), and OAR domain that is characteristic of this family of proteins [13]. Rx has been reported as an important regulator of vertebrate eye development. In *Xenopus*, Rx is expressed in the developing forebrain, retina, and pineal gland [112]. In addition, mouse embryos that carry a null Rx allele do not develop optic cups or eyes and also have a severely affected or absent forebrain [113]. Similar results have been observed in humans, mice, frogs, and medaka [112-118] that lack Rx expression. This data suggests that Rx is critical for the proper development of the vertebrate eye and forebrain.

Homeodomain genes typically encode transcription factors and indeed, Rx has been found to have the ability to regulate transcription [110]. Specifically, Rx can bind to and mediate transcription through the photoreceptor conserved element 1 (PCE-1) sequence that is found in all known photoreceptor cell-specific genes. Rx was shown to activate transcription of the mouse arrestin and interphotoreceptor retinoid-binding promoters which both contain a PCE-1 site in reporter assays [104]. Additionally, overexpression studies have indicated that Rx may have repressive functions because many neuronal differentiation markers are repressed when there is an abundance of Rx protein in the developing *Xenopus* eye and brain [119]. However, the same group also showed that an Rx-engrailed fusion protein acted as an antimorph, suggesting that Rx normally functions as an activator. It has also been suggested that all proteins that
regulate the rhodopsin promoter are transcriptional activators. Therefore, it is clear that there is some contradiction of the transcriptional functions of Rx and that more studies are necessary in this area.

The Rx protein contains a conserved octapeptide motif (eh-1 like motif) that has been shown to mediate repression in other members of the paired-type homeodomain protein family [81]. It is not known whether Rx can mediate repression via the octapeptide motif. Rx-like (RxL, QRX, RaxL) genes have been identified in *Xenopus*, chicken, and human. The gene product is similar to Rx in the homeodomain and OAR domain, however it lacks an octapeptide motif [120-122]. In *Xenopus*, chick, and humans it has been shown that Rx is a weak activator and that RxL, QRX, and RaxL can activate its target promoter to a greater extent than Rx, presumably because of the lack of an octapeptide [118-120].

Despite the findings that Rx can activate transcription through a PCE-1 site, little else is known about the molecular function of Rx as a transcription factor or what protein domains are important for its activity. Previous results in our lab indicate that Rx is a weak activator of the *Xenopus* rhodopsin promoter (XOP) despite the presence of an OP. It has been shown by EMSA and chromatin immunoprecipitation (ChIP) experiments that Rx can bind to the PCE-1 site within the XOP both *in vivo* and *in vitro* [104] (and our unpublished data). We used an XOP luciferase reporter construct to study the effects of mutations or deletions within the N-terminus and OP to assay Rx function and to determine the function of the OP. Our findings show that Rx is a weak activator, but that the Rx OP has a repressive function. This data also suggests that Rx has the ability to
confer repression via the OP since mutation or deletion of this domain results in an increase of activity. This work provides a better understanding of the results that have previously been reported on the transcriptional involvement of Rx in eye development.

4.3 Experimental Procedures

4.3.1 Cell Culture and Transfections

Cos 7 cells were maintained in Dulbecco’s modified Eagle medium that was supplemented with 10% Calf Bovine Serum, 100 units of the antibiotics penicillin/streptomycin, and 2mM L-glutamine at 37° at 5% CO₂. Sub-confluent cells were trypsinized and plated the day before transfection at a density of 5X10⁴ cells per well in a 24-well plate. Cells were transfected in triplicate with Fugene 6 transfection reagent (Roche) at a 3:1 DNA to Fugene ratio per manufacturer’s protocol with empty pCS2 plasmid included when necessary to bring up the total DNA to 400ng. Cells were harvested 24 hours later and lysates were assayed for reporter activity. A Dual Reporter Luciferase Assay Kit (Promega) was used in which a Renilla luciferase reporter construct under the control of the thymidine kinase promoter (pRL-TK, Promega) was co-transfected as an internal control. The internal control was transfected at a 1:20 ratio to the Xenopus rhodopsin luciferase reporter (XOP-luc) plasmid together with either no additional DNA, which served as the transcriptional baseline control, or DNA from wildtype or mutated constructs. The reporter assays were carried out as described per manufacturer’s protocol. A students’ t-test was performed to determine statistical significance of values.
4.3.2 Plasmids and Mutagenesis

Transfected DNA included pCS2 plasmids containing wildtype and mutant versions of the *Xenopus* Rx [113] coding sequence. *Xenopus* Rx was isolated from pSP64T vector with EcoRI and Bgl II (filled in) and sub-cloned into pCS2 with StuI. In addition, pCS2 plasmids containing the coding sequence for *Xenopus* Otx5b and XLmaf were also used. The L38P and the internal OP deletion constructs were made using Quick Change XL Mutagenesis Kit (Stratagene) using pCS2-Rx plasmid as the template.

L38P Forward: 5’ CTGCACAGTATAGAGGCCATCCCGGGGTTTGTGAAA3’
L38P Reverse: 3’ TTTCACAAACCCCGGGATGGCCTCTCTATGCAG 5’

-OP Forward: 5’ AACCCATCAAGGCTGGGGTTTGTGAAAAGAG 3’
-OP Reverse: 3’ CTCTTTCACAAACCCAGCCTTGATGGGTT 5’

The Rx N-31 and N-38 deletion constructs were made via PCR using pCS2-Rx as the template with the following primers:

N-31 Forward: 5’ GATCGAATTCAAGGCTGCTGAGGCC 3’
N-38 Forward: 5’ GATCGAATTCCGGGTTGTGAAAGAG 3’

Rx cds Reverse: 3’ GATCCTCGAGGTTTACCAAGGCTTGCC 5’

The PCR products were cloned into the pCR 2.1-TOPO vector (Invitrogen) and then isolated and sub-cloned into pCS2 using EcoRI and XhoI restriction sites. The Rx M10 start site mutant was made via PCR of the pCS2-Rx template using the following primer and the Rx cds reverse primer described above. The PCR product was cloned into the pCR 2.1-TOPO vector (Invitrogen) and then sub-cloned into pCS2 using EcoRI and XhoI restriction sites.
RxM10 Forward: 5’ GATCGAATTCTGAATGGTAGCTTCTCTCTGTCTG 3’

The Rx construct in which the M10 start site was mutated to an alanine was made using the Quick Change XL Mutagenesis Kit (Stratagene) using pCS2-Rx as the template with the following primers:

M10A Forward: 5’ CAGCCCTTCCCTGGCGGCTGATGG TAGC 3’
M10A Reverse: 3’ GCTACCATCAGCCGCCAGGAAGG GCTG 5’

The *Xenopus* opsin promoter construct used was described previously (Pan, 2006).

### 4.4 Results

#### 4.4.1 Rx is a weak activator despite the presence of an octapeptide motif

It has been reported that the eye transcription factors, Otx5b and XL-maf act to cooperatively activate the *Xenopus* rhodopsin promoter (XOP) primarily through the BAT and NRE sites [123]. The *Xenopus* rhodopsin promoter regulates transcription of the rhodopsin gene in rod photoreceptors and contains a conserved PCE-1 site. We have previously shown that Rx, together with Otx5b and XL-maf is a weak activator of the *Xenopus* opsin promoter in luciferase reporter assays (Fig.4.1). In fact, in our transfection assays with Cos 7 cells, Rx is only seen to have an affect on the XOP in the presence of these two factors.

Previous data in our lab with a closely related family member of the Rx protein, Arx, showed that it is a potent transcriptional repressor and the octapeptide (OP) is in part responsible for this observation (see Chapter 2). This finding led us to speculate that it is possible that Rx is a weak activator of transcription due to the presence of an octapeptide motif. Arx and Rx contain slightly different octapeptide sequences, however key
residues are conserved between the two proteins as well as between other homeodomain proteins (Figure 4.2). To determine if the Rx OP exhibits repression, we replaced the Arx OP with the Rx OP and assayed function in a luciferase reporter assay. In the context of Arx, the Rx OP has repressive function as compared to wildtype Arx (data not shown). This result confirmed that the Rx octapeptide has a repressive function, however in the context of Rx this function in not observed. Similar results were obtained when the Rx OP was added to the RxL protein, which lacks an OP, further confirming that the Rx OP can mediate repression. These results led to the speculation that a region in the Rx protein may be inhibiting repression by the OP.

4.4.2 N-terminal mutations and deletions affect the function of Rx as a weak activator

To study the OP further we made a version of Rx that deleted the OP as well as the N-terminus. We found that deleting the N-terminus and the OP resulted in lack of Rx function on XOP. We hypothesized that the N-terminal region just before the OP may possess a function that inhibits the repressive activity of the OP. We made a series of deletion and mutation constructs to test in the reporter assay. Constructs tested were: an internal OP deletion, deletion of the first 31 amino acids before the OP, deletion of the first 38 amino acids that included the OP, and a L38P point mutation that occurs at a highly conserved leucine residue in the OP (Fig. 4.3). We made this point mutation based on the previous evidence that this residue is important for repression via the OP in Arx (see Chapter 2). Interestingly, the OP deletion and the L38P point mutation exhibited a higher activity than wildtype, confirming that the Rx OP harbors a repressive function (Fig. 4.4). However, the results from the other deletion constructs were not as
clear. We hypothesized, based on our previous results, that the N-terminal portion of Rx may act to weaken the repressive effects of the OP and that it could be acting in this manner because it was important for activation. Deletion of the first 31 amino acids of Rx activated XOP as compared to wildtype and showed slightly less activity than the OP mutant constructs (Fig. 4.5). Deletion of the first 38 amino acids of Rx, including the OP, did not significantly affect the reporter activity as compared to wildtype Rx (Fig. 4.5). While it seems that the region N-terminal to the OP exhibits a function, additional experiments are necessary to understand how its molecular function plays a role in contributing to the activator or repressor functions of Rx.

4.4.3 Amino acids 1-9 of the N-terminus do not affect Rx function

To narrow down the amino acids in the N-terminus that are important for function, we took advantage of a previous report that found that Rx contains two ATG codons (M1 and M10) that are both translationally active start sites [124]. It was shown that less Rx protein was produced from the M1 start site however there was no phenotypic effect on Xenopus embryos that overexpressed this version of Rx, and the functional consequence of this finding was not examined [122]. We tested the functional significance of this finding by making a series of mutations within this region of Rx and testing them in a luciferase reporter assay. Three constructs were made: one that lacked the first nine amino acids, another in which the M10 was mutated to an alanine, and compared those to wildtype (Fig. 4.6). Altering the translation start of the Rx protein and removing the first ten amino acids of the N-terminus, did not affect Rx function on the opsin promoter, there was no difference as compared to wildtype (Fig. 4.7). Therefore,
the first nine amino acids in the N-terminus of Rx are not required for Rx function in this assay.

4.5 Discussion

The retinal homeobox gene (Rx) is essential for vertebrate eye development. Although there are numerous reports describing the effects that result from misexpression of Rx on eye development, very little is known about the functional characteristics of the protein. We made use of the fact that Rx can bind to and regulate transcription of PCE-1 sites [104] to study the effect of Rx mutant proteins on the transcription of the *Xenopus* opsin promoter (XOP).

Octapeptide (OP) motifs have been shown to mediate repression in other homeodomain proteins [81]. The Rx protein contains a conserved octapeptide motif, however in our studies Rx acts as a weak activator of XOP. In this study, we set out to determine whether the Rx octapeptide has an intrinsic repressive function and how mutation of this domain and the sequence N-terminal to it, affects the ability of Rx to activate transcription. We found that deletion of the OP or mutation of a highly conserved leucine (L38) within the OP results in an increase in activity. The L38P point mutation mimics that of the L33P mutation found in the Arx protein that is a cause of XLMR [55]. The L33P mutation in Arx results in decreased repression activity and a weakened interaction with the co-repressor Groucho [77] (see Chapter 2). Furthermore, the Rx OP can function to repress in the context of another paired-type homeodomain protein and transcriptional repressor, Arx. These data confirm that the Rx OP harbors a
repressive function. However in the context of the Rx protein, repression of XOP is not observed.

Previous data from our lab led us to speculate that the region just N-terminal to the OP may have a negative effect on its repressive function. We found that the N-31 deletion construct activated the XOP reporter as compared to wildtype. The N-38 deletion construct did not have an effect on activity. Our results do not provide clear evidence that this region is important to OP function, but suggests that the N-terminal portion of Rx may be involved in regulating its overall activity. Additional experiments are necessary such as identifying possible proteins that can bind this region to determine the mechanism in which it inhibits OP function. The Rx-L or Arx proteins may provide a useful tool in future studies of Rx function in domain-swapping experiments that may yield more informative data. Moreover, a cell type specific for eye or retinal tissue may provide clearer results than the use of Cos7 cells when it is considered that other retinal factors may be necessary for the normal activity of the XOP promoter. It is also possible that Rx function is not only context dependant, but also promoter dependant. Therefore, as new downstream targets are identified, the Rx mutants described here should be tested in a reporter assay with those targets. Additionally, identification of partner proteins that may cooperate with Rx on its targets would also be informative in deciphering Rx function as a weak activator versus a repressor.

Also interesting would be to determine whether co-repressors such as Groucho are able to modulate repression of Rx as seen with Arx (see Chapter 2). Both our lab and others [94] have shown that the Groucho co-repressors are strongly expressed in the
developing *Xenopus* eye and brain. It is possible that Rx acts as a repressor when it is expressed in similar developing regions as Groucho. It is also possible that a co-factor that binds to the N-terminus of Rx interferes with Groucho binding to the OP. Co-transfection experiments with the Rx constructs described in this work together with Groucho and different luciferase reporter constructs would provide information on whether the OP is modulated in a Groucho dependant manner.

These results suggest that not all OP-containing proteins are solely repressors. In addition, it suggests that there are mechanisms for regulation of the OP, possibly so that transcription factors can act as both activators and repressors. The ability to act as a bi-functional transcription factor would allow for the factor to act in a manner that is dependant on context. For instance, Rx may act as an activator in order to up-regulate transcription of genes necessary for early eye development. Conversely, in certain instances Rx may act as a repressor in order to repress genes involved in specifying certain cell fates. Alternatively, transcription factors may act as activators or repressors due to the context of the molecular environment in which they are expressed. For example, a transcription factor may be an activator at one stage of development due to the presence of co-activators that may not be present at other stages. Regulation of the OP may only occur when the factors that modulate its activity are present and therefore would enable Rx to be a repressor in some instances and an activator in others. The bi-functionallity of a transcription factor allows for diversity in the manner in which it regulates it downstream targets and allows for specificity of regulating different promoter
gene contexts. A molecular network of proteins act to coordinate transcriptional events that lead to the regulation of genes involved in the proper development of an organism. Elucidation of the molecular functions of Rx would provide a better understanding of how it regulates genes in both eye and brain development. In addition, the discovery of new downstream targets together with the functional dissection of the protein would aid in demonstrating the molecular characteristics of Rx.
Figure 4.1. Rx is a weak activator. A. Luciferase assay performed with lysates from Cos7 cells transfected with pCS2-Rx, pCS2-otx5b, and pCS2-Xlmaf with a XOP-luc reporter plasmid. Value is significant to Otx5b and Xlmaf alone. *p<0.05
Figure 4.2. Octapeptide Comparison. A. Arx and Rx octapeptide sequence comparison. Residues in bold are completely conserved. The conserved leucine is amino acid 33 in Arx and 38 in Rx.
Figure 4.3. Rx constructs. A. Schematic representation of N-terminal Rx mutation constructs used in reporter assays. Underlined portion represents the octapeptide motif.
Figure 4.4. Rx exhibits greater activity upon deletion or mutation of the OP. A. Luciferase assay performed with lysates from Cos7 cells transfected with pCS2-Rx, pCS2-Rx-OP, and pCS2-Rx L38P with or without the addition of Otx5b and Xlmaf co-activators together with a XOP-Luc reporter plasmid. Values are significant for Rx-OP and Rx L38P as compared to Rx upon the addition of Otx5b and Xlmaf. *p<0.05.
Figure 4.5. The N-terminus of Rx affects its function. A. Luciferase assay performed with lysates from Cos7 cells transfected with pCS2-Rx, Rx-OP, Rx L38P, Rx N-31, and Rx-38 with or without Otx5b and Xlmaf co-activators together with a XOP-Luc reporter plasmid. Values for Rx-OP, Rx L38P, and Rx N-31 are significant as compared to Rx upon the addition of Otx5b and Xlmaf. *p<0.05.
**Figure 4.6. Rx constructs.** A. Schematic representation of Rx translation start site (M) mutant constructs used in reporter assays. **OP:** octapeptide, **HD:** homeodomain, **OAR:** Ortho-Aristaless-Rx domain.
A. Luciferase assay performed with lysates from Cos7 cells transfected with pCS2-Rx, RxM10, or RxM10A with or without Otx5b and Xlmaf together with a XOP-Luc reporter plasmid. There was no significant difference between any of the values as compared to Rx.

**Figure 4.7. Amino acids 1-9 in the N-terminus do not affect Rx function.** A. Luciferase assay performed with lysates from Cos7 cells transfected with pCS2-Rx, RxM10, or RxM10A with or without Otx5b and Xlmaf together with a XOP-Luc reporter plasmid. There was no significant difference between any of the values as compared to Rx.
CHAPTER 5
DISCUSSION

5.1 Introduction

Homeobox genes are essential for the proper development of organisms. This work focuses on two homeobox genes that belong to the paired-type Aristaleless related family of homeodomain proteins. Both proteins are expressed in the developing nervous system and play important roles in regulating its development. Arx is vital to normal brain development and the proper migration of neurons whereas Rx is essential for the development of the vertebrate eye. Mutations in Arx are associated with X-linked mental retardation (XLMR) and other neurological pathologies. Mutations in Rx are rare however one mutation has been identified in anophthalmia (or lack of an eye) [118]. Considering the importance of these homeobox genes on development, the goal of this thesis was to provide new knowledge that contributes to the field of transcriptional regulation regarding these two factors. This work also provides insight into the differential manner in which the octapeptide (OP) is regulated in Arx (a strong repressor) and Rx (a weak activator) of transcription and the co-factors that are involved in mediating the repression. The different modes of OP regulation might be applicable to
other proteins of this family and may provide a way for proteins that contain an OP to possess both repressive and activation functions.

5.2 Results

5.2.1 Chapter 2

Chapter 2 of this thesis deals with the regulation of Arx as a transcriptional repressor. It was demonstrated that Arx is a transcriptional repressor with two independent repression domains: the octapeptide (OP) and a novel domain located at the C-terminus of the protein. In addition, it was found that disease-associated mutations affect the ability of Arx to act as a repressor. A point mutation located within the OP at a conserved leucine residue, which is mutated to a proline (L33P) in patients with XLMR, reduces the repression observed by Arx in reporter gene assays. In addition, it was found that the Groucho co-repressor enhanced repression of Arx and that the L33P or an OP deletion abolished this observation. This result agrees with previous results [77] in which it was shown that L33P weakened the interaction of Arx with the co-repressor Groucho.

Additionally, it was demonstrated that a C-terminal truncation that occurs at exon 5 of the protein and is found in patients with X-linked lissencephaly with abnormal genitalia (XLAG) also reduced the repression function of Arx in the reporter assay. It was further demonstrated that a region comprising that area (amino acids 398-448) of the C-terminus was indeed another repression domain (ORD) that was modulated by the co-repressor CtBP. Moreover, it was determined that Arx could bind to both CtBP and Groucho suggesting that transcriptional repression is mediated by two domains within the protein that are mediated by different co-repressors. This provides evidence that Arx
function can be modulated by different factors perhaps to regulate gene expression of downstream targets in a complex manner to allow for multiple configurations of the proteins in a context or time dependant manner.

5.2.2 Chapter 3

Chapter 3 of this thesis describes the potential DNA binding capabilities of Arx. Several attempts were made to identify a potential binding site that could be used in the reporter assays to test the affect of homeodomain mutations on transcription. Despite these attempts, a binding site was not identified. However, it can be deduced from the experiments that Arx does not bind to previously reported sites.

5.2.3 Chapter 4

In Chapter 4 of this thesis, the transcriptional activities of the retinal homeobox (Rx) gene were examined. Specifically, the potential transcriptional properties of the N-terminal portion of the protein were observed. It was found that Rx acts as a weak transcriptional activator of the *Xenopus* rhodopsin promoter (XOP). Rx also synergizes with two other retinal factors, Otx5B and XLmaf to activate XOP. In addition, it was shown that mutation or deletion of the OP renders it a better activator proving that the OP has a repressive function. Furthermore, it was found that the N-terminal region just before the OP might be involved in mediating the function of Rx either by inactivating OP activity or playing a role in activation. These potential functions for the N-terminus need to be explored further. Nevertheless, these results provide insight into the transcriptional function of Rx for which there is little known. This work also indicates
that Rx may act as a bifunctional transcription factor that may be dependant on the downstream target or other co-factors that are present.

5.3 Future Directions

This work of this dissertation has provided evidence for the transcriptional functions and regulation of two essential homeobox genes. It has shown a role for repression, activation, and modulation by other cofactors as important aspects of transcriptional control and how mutation that affects protein function may contribute to loss of regulation of downstream genes leading to disease. In addition, this work shows how two related proteins that each contains an OP are able to exert different transcriptional abilities. While contributing new knowledge to the field, this work has also raised many additional questions and potential areas for additional experimentation.

5.3.1 Chapter2

It was discovered in this work that Arx repression is modulated by two different co-repressor proteins each of which are involved in qualitatively different modes of repression: long-range vs. short-range. A transcriptional repressor that can repress by different mechanisms would allow for specificity in the type of repression necessary to occur at a given locus. For instance, repression of a gene that is needed for a temporary developmental event to occur may be regulated by a short-range repression mechanism. However, for more permanent repression that is necessary to silence several enhancers over a long distance that is to be maintained in the next cell generation, then long-range repression would better accomplish this. Another potential advantage of this type of modulation is to allow for tight regulation of repression that can occur at specific times.
during development. For example, Arx may recruit CtBP to promoters at one stage of
development and Groucho at another depending on the transcriptional event it is
regulating at that time, allowing for diversity in the types of processes Arx may be
regulating. For example in Chapter 2, the expression pattern for Arx and CtBP in
*Xenopus* embryos does not overlap until later stages in development. It would be
interesting to determine if Arx can modulate its repressive functions differently in early
versus late stages on specific downstream targets at that time in development.

In addition, it would be of interest to identify additional proteins that form the
different repressive complexes with CtBP and Groucho. This is be interesting in light of
the speculation that different repertoires of the repressosome complex, including histone
deacetylases (HDACs), may be recruited to different target genes [78]. A yeast-2-hybrid
(Y2H) screen or protein array analysis could be used to identify other potential protein
interactions. Also, the purification of the complexes at different developmental stages on
different promoters could be done together with mass spectroscopy to identify interacting
proteins.

Another area that could be followed up on is elucidating the three-dimensional
structure of the Arx protein. Computer program analysis could be performed to elucidate
the structure of the wildtype protein as well as mutated forms of the protein. For
instance, the point mutation (L33P) within the OP has deleterious effects on the
repressive functions of Arx and leads to disease. In fact the work in Chapter 2 shows that
the L33P mutation caused a comparable loss of repression to the OP deletion. This
suggests that L33 is essential to OP function. It is tempting to speculate that this is
because proline is an amino acid that induces bends in the protein and that this conformational change may cause alterations in or lack of protein-protein interactions of this region or other regions of the protein. Loss of protein-protein interaction would lead to misregulation of Arx, leading to aberrant regulation of its downstream targets, and therefore could result in disease.

In addition, it has been reported that truncations cause a more severe phenotype than missense point mutations [63]. Truncations that result in a non-functional or lack of protein due to an unstable or degraded transcript, would undoubtedly lead to disease. However, truncations that result in the deletion of domains important for the function of the protein could also lead to disease. This is of particular interest in light of the fact that this work shows that there is another repression domain located in the region where truncations occur that result in XLAG. This work also shows that Arx can interact with both CtBP and Groucho and that there may be some interplay between these two cofactors (see Chapter 2). It is possible that the C-terminus is involved in an intramolecular interaction with another portion of the protein or provides a favorable conformation for protein-protein interaction. This points to a possible function for the OAR domain in mediating protein-protein interactions since it is missing in C-terminal truncation mutations. A thorough examination of the three dimensional structure of the Arx protein to observe how the protein folds on itself and what faces are accessible for protein interaction in both the wildtype and mutants would help to provide answers to these questions.
The OAR (*Aristaless*) domain is present in all members of the *paired*-type *Aristaless* homeobox proteins, however its function is not well understood [12]. It was initially thought that this domain played a role in transcriptional activation of the Orthopedia protein [73]. More recently, the OAR domain has been suggested to have a function in attenuating transcriptional activation through intramolecular interaction with an activation domain in the N-terminal region of proteins such as PITX2, Prx1, and Cart I [13, 106, 125]. It was also proposed that in some contexts, the intramolecular interaction between the OAR and the N-terminus inhibited the DNA binding capabilities of the homeodomain [123]. In fact, recently it has been shown that the chromatin-associated high mobility group protein (HMG-17) which regulate chromatin structure and rates of transcription, can interact with PITX2 via its OAR domain [126]. Amen *et al* also found that HMG-17 could interact with the PITX2 HD, which resulted in an inactive form that could not bind DNA. However, in that report, it was not clear whether HMG-17 interaction with the OAR domain disrupted HD function. It was speculated that HMG-17 may play a role in directing HD-containing proteins to specific sites in chromatin and facilitate transcription by altering the chromatin structure [126]. Perhaps, if HMG-17 or another HMG protein could bind to the OAR domain of Arx, it might be involved in manipulating local chromatin structure in which deletion of the domain would result in a change in transcription. However, this scenario is likely more complex since deletion of the Arx OAR did not result in a change in transcriptional function (Chapter 2). It is possible that other regions in the C-terminus contribute to OAR function and deletion of just the OAR has no effect. Additionally, Arx reporter assays done with a true target
might yield different results (with an OAR deletion construct) than those obtained with an artificial reporter system since Arx may function differently on native targets. In any case, these recent reports provide examples of the manner (intramolecularly or intermolecularly by binding to other factors) in which the OAR may function to alter the transcriptional abilities of the proteins of which they are found.

The OAR domain may play similar roles in the function of both Arx and Rx. It is interesting to speculate that the OAR domain may play a role in the function of Arx since C-terminal truncations have such a deleterious affect on humans that carry such mutations. In addition, it is of interest for the work presented in Chapter 4 for the Rx protein in light of the fact that Rx may contain an N-terminal activation domain that could be masked by the OAR domain. Clearly, the conservation of the OAR domain among the Aristless group of proteins points to an important role for this domain, however additional studies are necessary to delineate the data that has been reported.

Mutations in Arx are known to cause disease and aberrations in the migration of neurons that is likely the primary cause of the brain malformations observed [50]. It would be of great clinical interest if the mutations in Arx could be further linked to the phenotype. For example, a Xenopus or zebrafish model could be used in which wildtype Arx protein expression is knocked down by morpholino, and a mutant version of the protein expressed in its absence. The brains of these animals could then be analyzed for structural malformations. A more detailed analysis of the brain and migrating neurons of the cortex could be performed using a mouse knock-in strategy in which mutant forms of
Arx could be analyzed for their affect on mouse brain development, which has been well characterized.

In addition to the proposed experiments, a more comprehensive mutational analysis should be performed on the Arx protein. There are additional regions of conservation outside of the octapeptide, homeodomain, and OAR domain. We tested versions of Arx in which other conserved regions such as the OAR domain were deleted, however these did not result in any change on transcriptional function. It is possible that these domains function in combination with other domains and that mutational analysis in which multiple regions were deleted is necessary. It is also possible that testing these versions as well as the ones previously tested (see Chapter 2) with an *in vivo* reporter instead of the Gal4-tethered system would yield more information on the function of these domains.

Transfections performed in a cell type of neuronal origin may also yield more informative results than those from the Cos7 cells. This strategy would provide conditions that are in a more *in vivo* context because other neuronal specific factors would be present in those cells. Differing results were obtained for some of the constructs tested in Cos7 cells as compared to embryos (see Chapter 2). These differences are probably due to the molecular nature of a single cell as compared to a whole embryo. Embryos provide a more heterogenous environment, whereas a cell suspension is made up of only one cell type. That data provides evidence that context is an important factor in deciphering the transcriptional function of Arx.
In addition, Arx contains four polyalanine tracts (PolyA). Humans and mouse share all of these conserved tracts, however *Xenopus* and zebrafish contain only the fourth tract [53, 55, 71]. It has been suggested that PolyA tracts play a role in repression, however recent findings have found that they may have different functions [53]. The work done in Chapter 2 supports that notion due to the finding that repression by *Xenopus* Arx was mediated through two motifs that were not associated with PolyA tracts.

PolyA tract expansions in the ARX protein are associated with various neurological disorders [53-55, 75, 127]. It was shown that Arx could induce cell dysfunction and death of transfected neurons by forming nuclear inclusions, which may be the mechanism that results in disease of humans with PolyA expansion mutations [75]. In a more recent study, it was shown that ARX protein aggregation was due to an increase in PolyA tract length and resulted in a shift from nuclear to cytoplasmic protein localization that may result in partial loss of function, contributing to disease [128]. In addition, PolyA tract expansions have been reported in other transcription factors that also result in disease suggesting that these regions play an important role in protein function [74]. Although it is not exactly clear how PolyA tracts affect transcription factor function, it is speculated that they might play a role maintaining the proper orientation of other domains in the protein, contribute to protein interaction, or DNA binding [74].

This work did not address the role of PolyA tracts on the function of the *Xenopus* Arx protein. One way that PolyA tracts could have been studied is to have introduced them into the *Xenopus* Arx protein to determine how they might affect function as
compared to wildtype in the reporter assays. Additionally, in the previous studies of ARX protein aggregation, only the effect of expansion of the PolyA tract was observed. It would be interesting to delete the PolyA tracts to determine if protein aggregates were still formed. Furthermore, it is intriguing to speculate the role PolyA tracts have on protein function and why it seems that for Arx, only higher vertebrates contain these domains. It remains a possibility that PolyA tracts in Arx confer a function that is necessary only in higher vertebrates or that during evolution they were introduced into the gene in higher organisms. It is also a possibility that the PolyA tracts do not have a function per se, but when expanded as is in the case of some diseases, that they are able to cause aberrant protein function.

5.3.2 Chapter 3

In Chapter 3 of this thesis, the identification of an Arx binding site was pursued to obtain a site that could be used in reporter assays to test the effect of disease-associated mutations within the homeodomain on DNA binding and transcription in a more in vivo situation. Alternate approaches to those used in Chapter 3 would be to make use of microarray technology. Potential downstream genes could be identified by gene expression microarray by comparing RNA recovered from embryo lysates in which Arx protein expression was overexpressed or down-regulated. Additionally, ChIP on Chip technology could be used in which Arx targets could be precipitated with Arx and then analyzed further by microarray to identify downstream genes. This would contribute significantly to the field, as there has not been a direct neural downstream target
identified for Arx. Furthermore, this would allow for Arx function to be assayed in a more informative manner.

It has been found that homeodomain proteins can heterodimerize with other homeodomain proteins to exert differential functions despite binding to similar DNA sequences [102]. An interesting notion to explore is the identification of putative partners that contribute to the binding capabilities of Arx. This could be achieved though a Y2H screen in which the Arx protein is used as bait or through the use of a protein array. Once partners are identified, they could be used to assay function of Arx in a reporter assay with identified targets in which different combinations of proteins may regulate different sets of genes.

Previous results [76] indicate that Arx might function as an activator as well as a repressor. Identification of partner proteins would also provide information on whether other proteins that interact with Arx contribute to this phenomenon. For example, if a protein found to interact with Arx was known to be an activator of transcription, it might provide clues as to whether Arx contributes to transcriptional activation via its partner protein. To determine what regions of Arx are important for interaction, mutant forms of the Arx protein, (like the exon 5 truncation described in Chapter 2), could be assayed by co-immunoprecipitation (Co-IP) analysis to determine which domain may be important for the interaction. Elucidating interacting proteins will provide insight into the genetic networks Arx is involved.

Very recently, a report has identified three direct neural targets of Arx: Ebf3, Lmo1, and Shox2 [111]. Using a genome wide microarray analysis of subpallia micro-
dissected from mice deficient in Arx, 84 total genes were differentially expressed in the null versus wildtype brains. The genes found are involved in cell/neuronal migration, axonogenesis, neuron morphogenesis during differentiation, and regulation of transcription. It was found that 57 were up-regulated and 27 down-regulated in the null mice as compared to wildtype, supporting previous evidence that Arx functions primarily as a transcriptional repressor, but may also have a function as an activator. This new report adds much needed knowledge to the field and provides insight into the genes and genetic networks that Arx is involved and will prompt future studies into determining how mutation in Arx leads to disease. For instance, further analysis of the promoters of the target genes described would indicate the other transcription factor binding sites that are near the Arx binding site. This information could lead to the discovery of proteins that could potentially interact with Arx to regulate transcription of that gene.

Fulp el al confirmed their findings of Arx targets by EMSA. Chapter 3 of this dissertation was dedicated to finding an Arx target and EMSAs were employed as one of the main strategies. We were not able to detect specific binding of Arx to sites similar to those in the report: TAATTA. This could be due to differences in the methods used. We used in vitro translated protein whereas Fulp et al used bacterially expressed GST-tagged Arx protein was used. In an in vitro translated protein lysate, there are other proteins in the lysate that could potentially bind to the site tested. Fulp et al also used only the homeodomain of Arx, whereas the EMSA experiments in Chapter 2 used the full-length protein. Despite the use of a bacterially expressed His-tagged version of the Arx HD in the SELEX experiments of Chapter 2, non-specific results were obtained. It is possible
that the protein used contained residual bacterial proteins that interfered with specific binding. Additionally, a flaw in the experiment or insufficient washing could have resulted in non-specific binding.

5.3.3 Chapter 4

In the last chapter of this dissertation, the transcriptional function of the retinal homeobox gene (Rx) was examined. It was found that the N-terminal portion of the Rx protein just before the OP might contain a functional motif that is involved in activation that inhibits repression through the OP. To further explore this, domain-swapping experiments could be performed. The N-terminal portion of Rx could be added to Arx, which shares similarity in protein sequence to Rx but does not contain this region in its N-terminus. This approach could assess whether this motif is involved in activation since Arx is a strong repressor it could be observed whether its function changes upon addition of this domain.

Another approach to help determine whether this domain is involved in activation would be to ask whether it binds to activator proteins. A protein screen could be performed to identify putative partners for Rx. Mutant versions of the Rx protein used in a Co-IP would help to determine which regions of the N-terminus are important for physical interaction. If an activator protein is identified it can then be used with the mutant versions of Rx in the reporter assay to further study the functional significance of that motif.

The work in Chapter 4 also indicates that the Rx OP is involved in repression since removal or mutation of this domain renders the protein more active as assayed on
the rhodopsin promoter. Rx is important for both the early development of the vertebrate forebrain as well as the eye [113]. Once additional downstream targets other than rhodopsin are identified for both the eye and brain, it would be interesting to assay Rx function on the targets in a reporter assay. This would allow for the determination of whether Rx can act as a repressor in some contexts and an activator in others during the development of eye and the brain. Identification of other downstream targets would also provide an additional tool for studying Rx function and interaction with partner proteins.

To further explore the repressive functions of the Rx OP, reporter assays similar to the ones described in Chapters 2 and 4 could be performed with Rx and Groucho. This would provide evidence as to whether the Rx OP is modulated by Groucho as shown for Arx (see Chapter 2). The OP mutant and deletion constructs could also be used in the assay to further delineate whether Groucho is able to mediate repression via the Rx OP and whether the conserved leucine at position 38 is important for that. In Chapter 2 the expression pattern for CtBP is shown to be in both the developing brain and eye. Although the C-terminal region of Rx is much different than that of Arx, it would be interesting to test whether this co-repressor could mediate repression by Rx.

The work in Chapter 4 demonstrates that the N-terminal portion of Rx may affect its function. In addition, Chapter 4 describes a repressive function for the OP of Rx despite the observation of Rx as an activator in reporter assays. This seems to be contradictory to the known function of the OP. Therefore, there must be a mechanism that renders the Rx OP inactive. Chapter 4 demonstrates that the N-terminal portion of Rx may affect the function of Rx. It is possible that this region is responsible for
recruiting activators thereby inhibiting the function of the OP. Additionally, it is also possible that the OAR domain plays a role in masking the OP through an intramolecular interaction so that co-repressors are not able to bind. Whatever the mechanism, this work raises the question as to how proteins with very similar domains are capable of exerting different transcriptional functions.

5.4 Concluding Remarks

The work contained in this dissertation demonstrates the transcriptional functions of the Aristaless-related homeobox proteins, Arx and Rx. Both of these genes are essential to the normal development of organisms. Consequently, mutations in Arx cause serious brain malformations that result in disease while lack of expression of Rx results in the loss of eye formation.

This work shows that despite the fact that both proteins contain an OP, they are able to exert different transcriptional functions. This finding suggests that regions outside of the OP must be important for the regulation of this domain. These other regions may be involved in protein-protein interactions that affect co-factor binding to the OP. In the case of Rx, the N-terminus just before the OP may be involved in this type of regulation. It is possible that it may bind an activator protein that inhibits or negates co-factor binding to the OP. As for Arx, it contains two repression domains and it is not clear how these domains could be affected such that activation would occur. However, it is possible that in certain instances of development repressive cofactors are unable to bind and activation could occur. One such factor that is postulated to antagonize Groucho in certain contexts, is a shorter member of the family that heterodimerizes with
the full-length Groucho therefore acting as a dominant negative form [81]. In addition, post-translational modifications could alter the binding of co-factors to the protein as described in Chapter 4. There are several potential explanations for the differences in OP function of these two proteins. This work provides a basis for the mechanisms involved, however further experimentation in this area is necessary to explain these possibilities.

The data in this dissertation provides details into the possible mechanisms of repression and OP regulation in two transcription factors of the *paired*-type *Aristaless* homeobox family. The results of this work in addition to future research will contribute knowledge to the body of previously published reports in this field describing the importance of these genes in transcriptional regulation and central nervous system development.
LIST OF REFERENCES


126


