INFORMATION TO USERS

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

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

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

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

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

UMI
University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700  800/521-0600
Studies of the TgHBox4 homeobox gene in *Tripneustes gratilla*

Vansant, Gordon, Ph.D.

University of Hawai'i, 1994
STUDIES OF THE TGHBOX4 HOMEOBOX GENE IN TRIPNEUSTES GRATILLA

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN BIOMEDICAL SCIENCES (GENETICS -CELL, MOLECULAR, AND NEUROSCIENCES) MAY 1994

By Gordon Vansant

Dissertation Committee

Tom Humphreys, Chairperson
David Haymer
Terrence Lyttle
Howard Mower
Eric Rosenthal
ACKNOWLEDGEMENTS

I would like to thank everyone who has helped me during my graduate studies but would like to specifically recognize certain individuals. First comes fellow graduate student and lab members. I owe much gratitude to Dr. "Dicky" Lum for being my mentor early on and teaching me what I needed to know to work independently. Thanks also to Lynn Godfrey for all her help and a shoulder to whine on. I would also like to thank and recognize other graduate students who have assisted me in many ways including Gordon Wang, Ruben Carlos, Brian Turano, and Ed Metz. I must give special thanks to Henning Ren for making me realize the true value of the PCR reaction and other intellectual insights. Now to the "professionals" of Kewalos. First I would like to thank Francis Okimoto and Geri Mitomi for making, deadlines, ordering, and other difficulties as pleasant and easy to handle as humanly possible. I would like to thank Rich Chock for making sure I could make it to work all these years with mechanical assistance way above and beyond the call of duty. I would also like to thank Rich and Steve Sakata for helping me whenever I needed it. Last but definitely not least in the "professional" category is Cheryl Phillipson. Her help with graphic and artistic techniques especially in the area of photography have been essential to the presentation of my thesis work. I would also like to thank certain members of the "intellectual establishment". This includes Dr. Greg Dolecki for putting in a good word for me initially, Dr. Asai for his intellectual
insight and occasional encouragement (your still here??), and Dr. Grace Tang
for her caring advice and encouragement. I would also like to thank other
faculty including Dr. Barbara Gibbons, Dr. Ian Gibbons, and Dr. Linda
Wordeman and especially my committee members for their guidance and
advice throughout the years. Special thanks to Eric Rosenthal for gallons of
coffee and hands on assistance. Thanks to Tom Humphreys for accepting me
into his lab and providing insight and direction at critical points during my
graduate studies. Thanks also to Dr. Rob Maxson for being my "sixth committee
member". His advice and collaborative efforts were invaluable. I would also like
to thank Dr. Neil Reimer for his many oligos that were prepared quickly and
other help he provided including amino acid content data of a peptide and
some selected DNA sequence data. Thank also to Dr. Gail Grabosky and Frank
Cipriano for being "sounding boards" for my ramblings during the completion of
my thesis work. I also need to recognize Dr. Robert Kane who built Kewalos
and was the director for most of my time here. He was very helpful and a
person I enjoyed conversing with. I also need to thank Dr. Fred Greenwood
and Dr. Marilyn Dunlap for supporting my research (financially) during most of
my graduate tenure.

Finally and most importantly I would like to thank my wife Melissa Garza-
Vansant for her support and patience, especially in the last two years during
which the end of my graduate schooling has been a moving target.
ABSTRACT

This study further characterized TgHBox4, an Abd-B class homeobox gene expressed as a 4.4kb and a 3.8kb mRNA in early and late embryogenesis, respectively, and as a 3.6kb mRNA in adult tissue of the sea urchin, Tripneustes gratilla. A full length cDNA clone was of 3623 base pairs encoding a 33,312kd protein was selected from an intestinal mRNA library. The cDNA clone was used to select a genomic clone containing a 5’ exon of 830bp, which along with an original 3’ exon of 2800bp accounted for the full cDNA sequence. Sequence analysis of the cDNA and the 5’ genomic clone indicates that the A at the start site of the cDNA may be the transcription start site, with a TATA box 30 nucleotides upstream. The first in frame ATG codon is 210 nucleotides downstream of this transcription start site but the ORF encoding the homeobox protein extends 5’ of the start site another 40 nucleotides including an upstream in frame ATG codon. The late embryonic message, which is slightly larger than the intestinal mRNA may be produced by use of an alternate upstream promoter.

A TgHBox4 sequence was expressed in bacteria and the protein used as an immunogen. Immunofluorescent studies using the antibodies localized TgHBox4 expression during development. The TgHBox4 proteins appear generally early in embryogenesis and become slowly restricted to the ectoderm cells over the spicule forming primary mesenchyme cells by gastrula stage. As gastrulation proceeds TgHBox4 proteins accumulate in the invaginating gut and became
progressively restricted to the region of the digestive tract by pluteus stage. The TgHBox4 bacterially expressed protein was also used in gel shift assays to study its binding to an enhancer of the late H2B histone gene of *S. purpuratus*. The anti-TgHBox4 antibodies specifically blocked this binding as well as the binding of the endogenous nuclear factor to the enhancer sequence, indicating that the TgHBox4 gene may regulate late embryonic histone synthesis. It is speculated that the protein product of the larger TgHBox4 message produced in early embryogenesis is involved in the determination of cell patterns which guide mesenchyme cell localization. The protein product of the smaller message produced later in embryogenesis may function to enhance H2B histone gene and other genes expression in dividing cells restricted to the gut in late embryos. The late expression of TgHBox4 in the vegetal plate structures derived from the most posterior parts of the blastula may relate to the posterior expression of the Abd-B class genes in the Hox gene cluster.
TABLE OF CONTENTS

<p>| ACKNOWLEDGEMENTS | .......................................................... iii |
| ABSTRACT | .......................................................... v |
| LIST OF FIGURES | .......................................................... ix |
| CHAPTER I. INTRODUCTION | ......................................................... 1 |
| CHAPTER II. MATERIALS AND METHODS | .................................................. 41 |
| DNA gels | ......................................................... 41 |
| Plasmid minipreps | .................................................... 42 |
| DNA precipitations | ..................................................... 43 |
| Restriction digests | ................................................... 44 |
| Ligations | ......................................................... 45 |
| Probe Production | ....................................................... 46 |
| PCR screening | ......................................................... 48 |
| cDNA and genomic DNA library screenings | ........................................... 49 |
| Southern Blots | ......................................................... 57 |
| DNA sequencing | ......................................................... 58 |
| Protein gels using mini-protean II dual slab gel | ........................................... 59 |
| Homeobox fusion protein expression construct | ............................................ 60 |
| Protein purification | .................................................. 63 |
| Production of antibodies | .................................................. 64 |
| Affinity purification of antibodies | .................................................. 65 |</p>
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preclearing of affinity purified antibodies</td>
<td>70</td>
</tr>
<tr>
<td>Western blots</td>
<td>71</td>
</tr>
<tr>
<td>Rearing of embryos</td>
<td>74</td>
</tr>
<tr>
<td>Harvest of embryos</td>
<td>75</td>
</tr>
<tr>
<td>Fixation, blocking, and immunofluorescent staining of embryos</td>
<td>75</td>
</tr>
<tr>
<td><strong>CHAPTER III. RESULTS</strong></td>
<td>77</td>
</tr>
<tr>
<td>Expression construct production and induction</td>
<td>77</td>
</tr>
<tr>
<td>Protein purification and antibody production</td>
<td>83</td>
</tr>
<tr>
<td>Immunofluorescent localization of TgHBox4 expression</td>
<td>88</td>
</tr>
<tr>
<td>Isolation and characterization of intestinal cDNA clone</td>
<td>91</td>
</tr>
<tr>
<td>Isolation and characterization of genomic clone</td>
<td>106</td>
</tr>
<tr>
<td><strong>CHAPTER IV. DISCUSSION</strong></td>
<td>117</td>
</tr>
<tr>
<td><strong>APPENDICES</strong></td>
<td>130</td>
</tr>
<tr>
<td>A. Development of the sea urchin embryo</td>
<td>130</td>
</tr>
<tr>
<td>B. Library of cTgHBox4-int cDNA sequencing reactions</td>
<td>136</td>
</tr>
<tr>
<td>C. Complete cTgHBox4 DNA sequence</td>
<td>141</td>
</tr>
<tr>
<td><strong>BIBLIOGRAPHY</strong></td>
<td>147</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>79</td>
</tr>
<tr>
<td>3</td>
<td>82</td>
</tr>
<tr>
<td>4</td>
<td>85</td>
</tr>
<tr>
<td>5</td>
<td>87</td>
</tr>
<tr>
<td>6</td>
<td>90</td>
</tr>
<tr>
<td>7</td>
<td>93</td>
</tr>
<tr>
<td>8</td>
<td>97</td>
</tr>
<tr>
<td>9</td>
<td>99</td>
</tr>
<tr>
<td>10</td>
<td>102</td>
</tr>
<tr>
<td>11</td>
<td>105</td>
</tr>
</tbody>
</table>

Figure 1: Homeotic protein DNA binding to the ATTA consensus sequence; DNA-protein contacts and three dimensional structure

Figure 2: TgHBox4 expression construct

Figure 3: Characterization of the bacterial expression protein

Figure 4: Characterization of the anti-TgHBox4 expressed protein antibodies. Reaction to *E. coli* lysate

Figure 5: Characterization of the anti-TgHBox4 expressed protein antibodies. Reaction to TgHBox4 peptide

Figure 6: Affinity purification of the anti-TgHBox4 expressed protein antibodies

Figure 7: Spatial localization of TgHBox4 proteins during embryogenesis

Figure 8: Alignment of homeobox containing restriction fragments of TgHBox4 intestinal cDNA and TgHBox4 genomic DNA

Figure 9: Translation of open reading frame of TgHBox4 intestinal cDNA

Figure 10: Genomic Southern blots reacted to cDNA intestinal probes

Figure 11: Alignment of the TgHBox4 5' homeobox containing genomic clone with the TgHBox4 cDNA intestinal clone
<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>108</td>
</tr>
<tr>
<td>Sizing of upstream exonic containing genomic TgHBox4 clone with PCR analysis</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>110</td>
</tr>
<tr>
<td>Alignment of homeobox containing restriction fragment of TgHBox4 intestinal cDNA with TgHBox4 genomic clone containing upstream exon DNA</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>114</td>
</tr>
<tr>
<td>5' region of upstream genomic clone gTgHBox4-2 including potential open reading frame</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>116</td>
</tr>
<tr>
<td>Alignment of TgHBox4 exons with TgHBox4 intestinal cDNA</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Embryogenesis is an intricate process, beginning with a relatively homogeneous single cell zygote and leading to a complex multicellular organism. During this process, cells must be produced and their developmental fate determined so that they differentiate into each specialized tissue type. A directive organizational plan must be established which provides the proper temporal coordination of gene expression such that all these specialized cell types develop in an orderly manner into a complete organism with fully functioning, coordinated tissue and organ systems.

It is not difficult to understand that complex processes such as embryogenesis will exhibit cases where correct organizational determination does not occur. One general type of developmental abnormality, defined by Bateson (1894) as homeosis, is the development of one body part into the likeness of another. Their genetic basis have been studied most extensively in Drosophila (E.B. Lewis, 1978). A complex of what appeared to be eight genes, which have been designated the HOM-C, is primarily responsible for such homeotic transformation. This includes the Bithorax gene cluster (BX-C), which controls the developmental pattern of thoracic and abdominal segments of the
fly. Mutations in this gene region produced homeotic transformations of the thoracic and, or abdominal structures of flies. Mutations in a related gene cluster, the (Antennapedia gene cluster (ANT-C))(Kaufman et al., 1980), produced homeotic transformations of structures in the head region of flies.

Lewis noted that these gene clusters have similar functions in different body parts, and suggested that they may be evolutionarily related, possibly arising from a primal gene that was duplicated, with the genes diverging as time passed. He thought the genes would have similar sequences and encode similar proteins. DNA sequences from selected genes, specifically the 3' regions of Antennapedia (Antp) and fushi tarazu (ftz) in the ANT-C, and the 3' region of the Ultrabithorax (Ubx) of the BX-C were found to cross hybridize (Scott and Weiner, 1984). These cross hybridizing regions were cloned, sequenced, and found to represent 180 base pairs in the 3' end of the open reading frames. These regions in Antp and Ubx are 79% similar in DNA sequence and encode peptides with 87% amino acid similarity (McGinnis et al., 1984). The higher amino acid sequence similarity between different homeoboxes relative to the DNA sequence similarity suggests that the conservation of sequence is due to an important functional aspect of the protein domains. Since these highly conserved DNA regions were found in genes whose mutation resulted in homeotic transformations, they were named homeotic sequences, or more commonly, homeoboxes (McGinnis et al., 1984). Since both BX-C and ANT-C genes have related sequences and, as described
later, are adjacent in many organisms, the genes were designated the HOM-C complex.

An increasing number of developmental genes of Drosophila have been found to include a homeobox. Homeobox sequences were mapped to the genes of the intra-abdominal region of the BX-C complex as well as to the region of the Antp locus corresponding to the proboscipedia (pb), zerknult (zen), and Deformed genes (Dfd) (McGinnis et al., 1984). Many other homeoboxes have been found in genes not associated with homeotic mutations, indicating a more general function for the homeobox. Homeobox containing genes were soon found in many other species. Cross hybridization of homeobox probes identified many in the genomes of chickens, humans, mice, and frogs (McGinnis et al., 1984), with clones of genes containing homeoboxes isolated from Xenopus laevis, mice, and human genomic and cDNA (Carrasco et al., 1984; McGinnis et al., 1984; Levine et al., 1984). Since the identification of homeobox containing genes in these organisms, their presence of has been detected in other organisms such as annelids, mollusca, echinoderms, and nematodes (Holland and Hogan, 1986; Way and Chalfie, 1988).

Comparative analysis of the protein "homeodomains" encoded by the three Drosophila homeoboxes and a Xenopus homeobox using the Dayhoff protein sequence bank revealed similarities to the yeast mating type regulatory proteins (Shepherd, J.C.W., 1984) and some bacterial regulatory proteins (Laughon and Scott, 1984). These yeast mating type regulatory proteins had previously been
shown to have regional homologies to prokaryotic gene regulatory proteins implicated in DNA binding (Pabo and Sauer, 1984).

These prokaryotic proteins have been shown through crystallographic studies to have conserved alpha-helical structures that computer modeling studies indicate are the DNA-binding sites (Anderson et al., 1981; McKay et al., 1981; Pabo et al., 1982; Matthews et al., 1982; Sauer et al., 1982; Weber et al., 1982). Binding of DNA by these proteins is supported by chemical modification studies that show specific operator DNA sites are protected when bound by regulatory proteins (Pabo and Sauer, 1984).

The DNA binding structure of the prokaryotic proteins is a two helix motif with one helix binding the edge of the bases in the major groove of the DNA binding site with the second helix laying across the first helix and contacting the phosphates in the DNA backbone, holding the protein in position. The two helices are connected by a beta turn (Pabo and Sauer, 1984). Residues important for DNA binding by the first helix, as well as some residues of the second helix, and the beta turn, are shared among the prokaryotic proteins and the homeodomain containing proteins (Laughon and Scott, 1984).

The question of specific DNA binding by homeodomains has been pursued most extensively with Drosophila. An expression construct was made from entire coding sequences of the Drosophila homeobox containing gene engrailed (en), to produce a beta-galactosidase fusion protein. The protein displayed selective DNA binding activity to regions upstream of the open
reading frame of the en gene at relatively high salt concentrations which is characteristic of sequence-specific recognition (Desplan et al., 1985). It bound to upstream regions of the fushi tarazu (ftz) gene under similar conditions (Desplan et al., 1985). Using DNAase protection Hoey and Levine, (1988), showed that fusion proteins with homeodomains encoded for by the genes eve, en, zen, and paired(prd), could bind similar sequences in a 5’ region of the eve and en genes. By varying protein concentrations in binding assays, they discovered that all the fusion proteins except the eve protein, had a binding preference for the 5’ region of the en gene, suggesting competition among proteins for regulatory sites. Directed mutagenesis of eve protein encoding DNA was carried out to determine the role of various amino acids of the homeodomains. Mutant eve construct products lacking a central part of the homeodomain, including 6 of 8 of the residues that make up the putative binding helix, no longer bound their previously characterized binding sites.

Several studies investigated the role of sequences of the homeodomains in the activity of the homeodomain containing proteins and in the specific recognition of different DNA binding sites. Fusion constructs were made from the prd gene homeobox, including two with the encoded amino acid at position 9 mutated from proline, to either a glutamine, as in the ftz homeodomain, or to a serine as in the bcd homeodomain. The unaltered construct recognized and bound prd DNA binding sites, whereas the glutamine and serine mutant forms bound ftz and bcd sites, respectively (Treisman et al. 1989). Levine’s group
(Hoey et al., 1988) carried out binding experiments using full length fusion proteins with swapped homeodomains. An en protein with a eve homeodomain and an eve protein with an en homeodomain displayed binding preferences representative of the homeodomains they contained.

Evidence of sequence specific DNA binding activity of vertebrate homeodomains has also accumulated. A fusion protein containing the murine Hox 1.5 homeodomain bound a DNA region 5' of the Hox 1.5 gene. A protein in an embryonic nuclear extract that should contain the protein encoded by the Hox 1.5 gene, based on RNA Northern transfers, also bound to the same DNA region (Fainsod et al., 1986). A murine Hox 1.3 fusion protein produced in the eukaryotic baculovirus system was shown by DNAasel protection assays to recognize a region 5' of the Hox 1.3 coding region. A specific DNA region with an ATTA consensus was needed for binding (Odenwald et al., 1989). Fusion constructs producing human oct-2 homeodomain containing proteins, which normally bind the immunoglobulin octamer DNA site, showed that certain amino acid changes in one of the helix regions of the helix turn helix motif eliminated recognition of the octamer site (Ko et al., 1988).

Physical analysis of the homeodomain and its binding activity provided evidence the homeodomain containing proteins are DNA binding proteins. An Antp bacterial expression vector was constructed which contained just the homeodomain coding region of the Antp gene. The resultant Antp fusion protein was purified and analyzed using ultracentrifugation, and PAGE.
electrophoresis, and was also used in a DNA binding experiment. Based on the first set of analyses, they predicted the homeodomain is a monomer in solution. From DNAase I protection analysis and gel retardation assays of a promoter region of the ftz gene bound by the purified protein, it was determined that the homeodomain alone was sufficient for DNA binding (Muller et al., 1988).

NMR analysis was used to determine secondary structure of the purified Antp protein. Nuclear Overhauser Effects (NOE), with analysis of slowly exchanging amide proteins, was the focus of the first NMR experiment. This data predicted three helices from residue 10 to 52 (Otting et al., 1988). Analysis of the NMR data with computer modeling programs predicted a helix turn helix motif in the homeodomain similar to the ones found in many prokaryotic DNA binding proteins (Otting et al., 1988). Closer scrutiny of the purified Antp protein with more NMR spectroscopy led to the discovery of a total of 3 helices, with one being more flexible than the other two. The additional helices, and the predicted spatial arrangement of the one more flexible helix, are substantial differences in comparison to the prokaryotic proteins with homologous helical structures (Qian et al., 1989).

Mobility shift assays were used to determine the dissociation rates of a purified Antp fusion protein to synthetic produced oligodeoxynucleotides. Two DNA binding sites were created. One was modeled from sequences in the en promoter region, and one was modeled from sequences of the ftz enhancer.
Both have previously been shown to be bound by Antp class homeoproteins (Muller et al., 1988; Pick et al., 1990). The Antp fusion protein was found to bind to the sites as a monomer with high affinity ($K_d \leq 1.6 \times 10^{-9}$) in what kinetics studies characterize as very stable (1/2 life of \(~90\) minutes) (Affolter et al., 1990). These binding characteristics are in strong contrast to the less stable (1/2 life of a few seconds or less), lower affinity ($K_d \leq 10^{-5}$ to $10^{-6}$) binding states of prokaryotic DNA binding proteins (Ogata and Gilbert, 1978; Fried and Crothers, 1981; Pabo and Lewis, 1982; Hollis et al., 1988).

Interference studies of the binding reactions using ethylation and methylation of the binding sites determined that the homeoprotein docks its DNA binding sites on one side with its region of interaction across 3 phosphate backbones (Affolter et al., 1990).

A more complete analysis was done on the physical interaction of one of the Antp DNA binding sites and a bacterially expressed Antp homeodomain by using NMR analysis of the complex of the two structures. It was determined that the protein and DNA both have similar conformations when they are free or in the complex (Otting et al., 1990). The identity of many of the intermolecular 1H-1H Overhauser effects (NOE) was determined to calculate the docking conformation of the protein to an idealized B DNA model using an ellipsoid algorithm (Billeter et al., 1987a, 1987b). This assigned an apparent interaction between the third helix with the major groove of the DNA and additional interaction of the DNA with a polypeptide NH$_2$ terminal extension in front of the
helix turn helix motif. This unique interaction includes the arginine at position 5 fitting into the minor groove, which was the first indication of the spatial arrangement of the homeodomain's flexible N-terminal region. The DNA-homeodomain complex appeared similar to the prokaryotic DNA binding proteins (e.g. the 434 repressor) DNA-protein complexes, except for the interaction of the NH₂ terminal with the minor groove (Agarwal et al., 1988; Pabo et al., 1990; Otting et al., 1990).

Analysis of interactions of wild-type homeodomain fusions and various mutant fusions of the ftz protein with the same DNA binding sites as those used with the Antp protein were carried out. This led to a model of protein-DNA interaction that was very similar to the one that was predicted from the NMR analysis of the Antp protein DNA complex (Otting et al., 1990; Percival-Smith et al., 1990).

X-ray crystallographic structure analysis of the en homeoprotein bound to a DNA fragment (Liu et al., 1990) elucidated a model of the complex that was consistent with the one determined from NMR analysis of the Antp protein DNA complex. This included the determination that residue Ile-47 would make contact with the same base within the ATTA conserved region of the bound sequences, that Tyr-25 would interact with the same region of the phosphate backbone and that Arg-5 on the NH₂ flexible extension would also fit into the minor groove of the DNA (see Figure 1) (Otting et al., 1990; Kissinger et al., 1990).
Figure 1. Homeotic protein DNA binding to the ATTA consensus sequence; DNA-protein contacts and three dimensional structure. Figure 1A is a sketch displaying all the DNA protein contacts. Balls represent phosphates and the lettered triangles represent bases. Arrows point from the amino residues to the specific part of the DNA that they interact with. Figure 1B is a cartoon representation of the three dimensional arrangement of the homeodomain to the TATA target sequence. Contacts considered crucial for binding are included in this binding model.
Functional transcription factors must not only bind DNA specifically but they must also affect gene expression upon binding. The number of in vitro and in vivo binding studies, as well as research of gene interaction, primarily among Drosophila homeobox genes, is accumulating that strengthen the idea that homeobox genes function as transcription factors which bind promoter or enhancer elements in the DNA and alter the expression of genes.

Extensive studies have been made into the ability of the products of various homeobox genes to transactivate reporter constructs containing sequence elements that these gene products bind. Expression constructs were made from cDNA's encoding Ubx, Abd-B, and OTF-2, and their product's putative function characterized using four globin reporter plasmids. One reporter included a Ubx binding site (Beachy et al., 1988), one included the sequence TCAATTTAAAT, which has been shown to be a consensus sequence for homeodomain binding (reviewed by Levine and Hoey, 1988; Otting et al., 1990; Kissinger et al., 1990), and the other two included either the TF-1 or the TF-2 mammalian transcription factor target sequences. Expression in mammalian HeLa cells cotransfected with all possible combinations of expression vector and reporter plasmid analyzed with SP6-RNase analysis showed that Ubx and Abd-B both activated all four reporter plasmids, with the highest activation initiated by the Abd-B protein (Thali et al., 1988). Comparison of the binding sequences revealed a consensus region within all four of the binding domains, with at least 8 of the 10 bases being completely conserved. It
is understandable based on this that the OTF-2 gene product not only bound it's target site, but the Drosophila target sites as well (Thali et al., 1988). Thus on a functional level, the three homeodomain containing proteins are equivalent in that they all are capable of activating transcription by binding the same target sequences.

Another group examined the interaction between the protein encoded for by the bicoid (bcd) gene, and the promoter elements of the hunchback (hb) gene. The bicoid protein is a homeodomain containing maternal protein that is involved in specifying the anterior regionality along the anteroposterior axis of the Drosophila embryo, and has been found to be necessary for development of the head and thorax (Frohnhofer and Nusslein-Volhard, 1986). Bicoid is distributed as a gradient along the embryo, with it's highest concentrations being detected in the anterior tip (Driever and Nusslein-Volhard, 1988). The hunchback gene is in the gap gene class, and it's mutants have deletions of gnathal and thoracic segments (Lehmann and Nusslein-Volhard, 1987). It's anterior localized transcripts are not expressed in bcd mutants (Tautz, 1988), which suggests it may be regulated by bcd. Driever and Nusslein-Volhard expressed a full length bcd fusion protein that specifically bound to specific regions of the hb promoter. They also made chloramphenicol acetyltransferase (CAT) constructs, fusing the bcd binding sites of the hb promoter to the CAT gene and injected them into embryos derived from wild type and bcd mutant females. They found that the CAT gene was expressed at levels up to 50 times
higher in wild type derived embryos compared to the *bcd* mutants (Driever and Nusslein-Volhard, 1989).

Kuziora and McGinnis utilized homeobox swapping to produce a fly strain with a *Deformed* (*Dfd*) gene with its homeodomain encoding region replaced by the homeodomain encoding region from the *Ubx* gene and with a heat shock inducible promoter. The construct was used to produce a fly strain by P-element transformation. The resultant fly strains were heat shocked at the cellular blastoderm stage, and the embryos were fixed and examined within 3 to 5 hours by immunostaining. In contrast to a strain with a control *Dfd* gene under the control of a heat shock promoter, the chimeric protein did not activate the chromosomal *Dfd* gene to produce the *Dfd* protein (Kuziori and McGinnis, 1989). RNA in situ hybridization using an Antp probe and the resultant phenotypes of the transformants revealed activation of *Antp*, as would be expected under *Ubx* regulation (Kuziora and McGinnis, 1989).

How do similar homeodomain containing proteins specifically bind and differentially regulate genes? Some of the binding studies discussed previously suggest that the homeodomain’s primary structure contributes to specificity (Treisman et al., 1989; Hoey et al., 1988; Kuziori and McGinnis, 1989). Studies which show that different homeodomain containing proteins recognizing and binding the same DNA site (Thali et al., 1988; Hoey and Levine, 1988) make it clear that different homeodomains recognizing different binding sites is probably not the exclusive basis for specific gene regulation *in vivo*.
The research showing that homeoproteins bind to DNA and affect gene expression provide insight into how homeodomain proteins may differentially regulate genes by interacting as they recognize the same binding sites. One possibility is that an array of homeotic proteins may have a different affect on gene expression depending on which homeotic proteins are present. Jaynes and O'Farrell made reporter constructs by fusing six copies of an enhancer region of the ftz gene to the CAT gene containing plasmid and made expression constructs, fusing the coding sequences of the ftz gene or the en gene into an actin 5C promoter polyadenylation vector. ftz and en homeoproteins have been found to bind the ftz enhancer sequence in vitro (Desplan et al., 1985; Jaynes and O'Farrell, 1988). They cotransfected Schneider Drosophila cells with the ftz expression construct and the reporter construct and found ftz to activate the reporter gene when the ftz binding sequence was present in the CAT construct (Jaynes and O'Farrell, 1988). Cotransfection of Drosophila cells with an en expression construct and the reporter construct displayed a failure of the en gene product to activate the reporter gene. Upon cotransfection of Drosophila cells with both the en and ftz expression constructs, along with the reporter plasmid, the presence of the en protein repressed the ability of the ftz gene to activate the reporter gene (Jaynes and O'Farrell, 1988).

Cooperative and synergistic interactions, as well as repressive interactions among homeotic proteins, may play a role in the specificity of homeotic gene
regulation. Cotransfections of Drosophila culture cells with a CAT reporter plasmid containing the \textit{en} promoter region and expression constructs coding various homeotic proteins support this possibility. Cotransfections including expression constructs encoding the \textit{zen}, \textit{z2}, \textit{ftz}, or \textit{prd} proteins resulted in activation of an \textit{en} promoter reporter gene. Cotransfections with expression constructs encoding for \textit{eve} or \textit{en} failed to activate the reporter gene. It was determined that the \textit{eve} and \textit{en} gene products actually repressed the activation activity of the other gene products, and that a combination of two proteins like \textit{ftz} with \textit{prd} often synergistically activated the reporter gene (Han et al., 1989).

A homeodomain protein may interact with other transcription factors in a specific way which could be reflected in their regulatory activity. Four beta galactosidase expression constructs were made for transforming \textit{Drosophila} embryos. One included the \textit{en} consensus binding site, one the \textit{ftz} promoter, one the \textit{en} promoter, and one the heat shock protein 70 (hsp70) promoter. These were introduced separately into the \textit{Drosophila} genome using P-element mediated transformation and the resultant developing embryos were fixed and analyzed for beta-galactosidase expression. The constructs containing the \textit{ftz} and the \textit{en} promoter were not expressed in the embryo, and no beta-galactosidase expression was seen during the developmental period when \textit{ftz} is expressed. This is counter to the results obtained in transfection assays using cell culture (Vincent et al., 1990; Jaynes and O'Farrell, 1988). The heat shock promoter construct was specifically expressed in glial cells. The paradox of a
construct with a common binding element being expressed in a specific set of cells can be resolved if it is postulated that an additional protein was involved. It was proposed that some accessory protein interacts with the hsp70 site, and both this interaction and homeoprotein's interaction with it's recognition sequence is necessary for expression of the gene being bound (Vincent et al., 1990). Lack of expression of the other constructs and lack of ftz activation in the intact flies is probably a reflection of a more complex network of regulatory proteins found in the embryo that are absent in tissue culture cells.

Closer analysis of the bicoid protein and it's DNA binding targets have revealed how different binding affinities probably play a role in determining what genes a homeodomain containing protein will regulate. As discussed previously the hunchback gene has a DNA region upstream of it's transcription start site that is recognized and bound by the bicoid protein (Driever and Nusslein-Volhard, 1989). This region can be broken down into 6 sites, three that bind the bicoid protein with a relatively high affinity, and three that bind it with a relatively low affinity. These workers made constructs driven by heat shock promotors and expressing Beta-galactosidase fusion proteins. The constructs include a series with successive deletions of the binding sites, one which included the three high affinity binding sites and one which included the three low affinity binding sites. These constructs were introduced into Drosophila using P-element mediated germ line transformation, and their expression was analyzed in early Drosophila embryos. The trend in the pattern
of expression of the constructs with successive deletions ranged from the
normal hunchback pattern of expression, to patterns with expression more and
more limited to the anterior region of the embryo, with each deletion of a
binding site. The constructs with the high affinity binding sites displayed a
pattern of expression in the anterior half of the embryo like the normal
expression pattern of the hunchback gene, whereas the construct with the low
affinity binding sites displayed an expression pattern that was limited to the
anterior half of the normal area of hunchback expression, where the bicoid
protein is at it's highest concentration (Driever et al., 1989). Promoter sites like
this allow a protein's graded distribution to be translated into sharp regulatory
boundaries.

Also, other homeoproteins recognizing the same binding sites might
compete more effectively for the regulatory sites in areas where the graded
proteins concentration is low and repress the gene, resulting in even more
sharply defined spatial domains of gene expression. An interaction like this
could be envisioned with a homeobox gene like caudal, who's early expression
leads to a gradient that is the opposite of bicoid, with it's highest concentration
in the posterior of the embryo (Macdonald and Struhl, 1986).

Another group reported binding activities that were dependent on the
genomic environment. Zink and associates transformed Drosophila with beta-
galactosidase constructs that contained the Antp P1 and P2 promoter
fragments. Using immunocytochemistry they detected the polycomb (Pc)
protein binding these sites when they were incorporated into certain regions of polytene chromosomes, even though similar binding was not seen in vitro (Zink et al., 1991).

An additional possibility for the regulation of the binding specificity of homeodomains is posttranslational modifications. It has been shown that a homeobox protein is phosphorylated in vivo (Odenwald et al., 1989), and theoretically such modification could modify DNA binding activity in a temporal, and or spatial fashion. To summarize, there are several ways for the homeotic genes, which are expressed in a temporally and spatially restricted fashion, to interact in a coordinated manner to regulate the development of complex organisms starting from a fertilized egg.

If homeobox containing genes encode DNA binding transcription factors functioning during embryogenesis, their localization of expression and interactions must be characteristic of this function. Transcription and protein expression analysis has revealed that each of the HOM-C homeobox genes is expressed in a temporally and spatially regulated fashion. Genetic analysis shows that lack of any of the genes results in developmental transformations in areas where the gene would normally be expressed (reviewed by Akam, 1987). Specifically, absence of a homeotic gene product in a region results in the region taking on a more anterior identity.

Genetic and molecular analysis has led to an extensive elucidation of the overall regulatory cascade of genes responsible for the correct development of
the *Drosophila* embryo (for complete review see Akam, 1987). Genes involved in *Drosophila* development are divided up into two major classes, maternal genes and zygotic genes. The maternal genes are expressed before fertilization and provide the embryo with its first, albeit crude, anteroposterior specification. The zygotic genes interact with the maternal gene products to further refine this specification. The zygotic genes are divided up into three classes: the gap genes, the segmentation genes, and the homeotic genes. The gap genes and the segmentation genes act to define the correct number of repeating segments. The homeotic genes of the HOM-C give each segment its specific identity. The homeotic genes seem to specify cell fates by positively and negatively affecting the expression of each other, and of downstream effector genes in the regions where they are expressed.

Genes that directly affect the expression of other genes would be expected to have their protein products localized to nuclei. All immunolocalization studies to date have determined that the protein products produced by homeobox containing genes, including *Ubx*, *en*, *ftz*, *Antp*, and *cad*, are localized to nuclei (White and Wilcox, 1984; Beachy et al., 1985; DiNardo et al., 1985; Carroll and Scott, 1985; Carroll et al., 1986; MacDonald and Struhl, 1986).

Another interesting aspect of homeotic genes is their colinearity of expression. The genes in the 5' region of the clusters are expressed in more posterior domains of expression while the more 3' genes of the complexes are expressed in more anterior domains (Lewis et al., 1978). In other words, the
genes are arranged on the chromosome in the same linear order as they are expressed along the anteroposterior axis of the developing embryo.

After envisioning how these proteins carry out their regulatory role in Drosophila, the question that remains unanswered is what are the results of these regulatory actions in other organisms. What are the ultimate physiological and morphological results of the action of homeobox containing proteins in vertebrates? Do the homeobox genes interact to affect cell determination and pattern formation in other organisms, as they apparently do in the arthropod, Drosophila?

Originally, it was thought homeotic complexes arose by duplication and diverging events that occurred with the appearance of segment specialization during the evolution of the modern insect. The ancestors of insects were probably worm-like organisms with a head and a body consisting of repeating, rather similar segments. Two genes from an original homeotic gene that had duplicated would specify the difference in the development of the head and body segments. As the head and trunk segments specialized, the genes responsible for these determinations would reflect the changes. More duplication, and subsequent diverging events of the original homeotic genes, would allow for the genes to regulate the formation of more specialized segments.

Homeobox genes in vertebrates have a similar organization to that of Drosophila genes. However, in the vertebrates there are four similar
complexes. In the mouse they are termed Hox-1, Hox-2, Hox-3, and Hox-5 and they are located on chromosome 6, 11, 15, and 2 respectively (Bucan et al., 1986; Hart et al., 1985; Breier et al., 1988; Featherstone et al., 1988). The linear organization of the genes of each complex allow alignment of each subfamily in the different clusters and the different HOM-C Drosophila genes, suggesting the four clusters are the result of chromosomal duplication events during evolution (Hart et al., 1987; Duboule et al., 1989, Graham et al., 1988). Expression analysis of the murine clusters has shown that they display colinearity of expression in the same manner as the genes of Drosophila (Duboule and Dolle, 1989; Graham et al., 1989). Sequence comparisons between the genes of the different murine clusters, and between the murine genes and the genes of the Drosophila clusters, has revealed an interrelatedness between genes, that allows them to be categorized into subfamilies according to sequence similarities. Other organism's homeobox genes are being found to be in similar clusters, including clusters found in chickens, beetles, and humans (Wedden et al., 1989; Beeman, 1987; and Boncinelli 1988). In all these other organisms the representatives of the ANT-C and BX-C genes of Drosophila are not separated, but are found in one chromosomal cluster.

Speculating on the role of homeobox genes in vertebrate development based on the genes' functions in Drosophila, is not straightforward. Vertebrates appear to have some segmented features, since the skeleton and muscles of the organisms originate from an array of primary repeating embryonic units, the
somites (Harvey and Melton, 1988). But the fact that vertebrates and arthropods have homeotic clusters with very similar characteristics, implies these complexes were formed by duplication and divergence, before the evolution of the insect, from a simple worm like form with repeating segments to the modern form with specialized segments. The gene complexes evolved in a Precambrian common ancestor of arthropods and vertebrates which was probably not segmented (Kaufman et al., 1990; Hogan et al., 1985). Whatever role HOM-C genes play in *Drosophila* segmentation must have evolved subsequent to the formation of the cluster. Also the discovery of homeoboxes in the sea urchin, the topic of this thesis, which are not segmented, further confirms that the HOM-C homeobox genes evolved before segmentation in either arthropods or vertebrates (Dolecki et al., 1986).

Conservation of the arrangement and expression patterns of these genes in two phylogenetically distant animals suggests a functional importance of both of these characteristics. Some research has focused on the function of individual HOM-C homeobox genes, looking for any functional correlations in *Drosophila*, nematodes, mice, and frogs.

Studies with the nematode invertebrate *c. elegans* offer correlations between this organism and *Drosophila* that strengthen the argument that the HOM-C homeobox genes of each are involved in pattern formation through cell fate determination. In worms mutant at mab-5, which is an *Antp* class homeobox gene expressed in the posterior of the developing postembryonic organism,
undergo a homeotic transformation, taking on the fate of more anterior cell homologs (Kenyon, 1986). Transcription analysis reveals that the expression of mab-5 is in the regions where cellular transformation occurs in mab-5 mutants (Costa et al., 1988). It has been shown that mab-5 is involved in the direction of movement of migrating cells in the developing organism (Salser and Kenyon, 1992). Another C. elegans homeobox gene, egl-5, has been implicated by genetic analysis in the determination of the fates of cells in the tail region (Chisolm, 1991).

Transgenic mice studies are being used to determine the function of some murine Hox genes. The Hox-1.4 gene appears to have a role in the development of the gut. An additional Hox-1.4 gene in gene constructs including the 5′ region up to the adjacent homeobox gene Hox-1.3 was microinjected into the pronuclei of fertilized eggs. Positive weanlings were detected by analysis of DNA of tail cells with a transgene specific probe (Brinster et al., 1985). The positive mice had DNA containing SV-40 sequences present in the recombinant constructs. The introduced genes were found to have the same temporal and spatial expression pattern as the endogenous Hox-1.4 gene. Overexpression of this gene in the embryonic gut resulted in a megacolon phenotype (Wolgemuth et al., 1989).

The Hox-1.5 gene seems to be responsible for the correct development of many cell types. Mutated forms of the Hox-1.5 gene were used for gene targeting by homologous recombination (Capecchi, 1989; Capecchi, 1989b) in
pluripotent mouse embryonic derived stem cells. The Hox-1.5 minus/Hox-1.5 plus mice that were produced were crossed to produce mutant homozygotes (Chisaka and Cappechi, 1991). The homozygotes were athymic, aparathyroid, and had defects in the throat, heart, arteries, as well as craniofacial abnormalities (Chisaka and Cappechi, 1991).

Transgenic mice lacking a functional Hox-1.6 gene were produced by a similar approach. Resultant transgenic mice displayed abnormal inner ears and bones of the skull, delayed hindbrain closure, missing cranial nerves, and missing ganglia in the region of rhobomeres 4-7 (Lufkin et al., 1991). Rhobomeres are transient periodic bulges in the hindbrain that reflect the segmental basis of development of the neuronal pattern (Lumsden and Keynes, 1989).

Hox-1.1 overexpression seems to disrupt ectoderm and mesoderm differentiation at the time endogenous Hox-1.1 is expressed (Mahon et al., 1988; Castile et al., 1990). The Hox-1.1 gene, driven by the chicken beta-actin promoter, was transfected into mice. Embryos that had Hox-1.1 ectopically expressed are born with craniofacial abnormalities and malformations of the basioccipital bone, the atlas, and the axis, including an additional vertebra similar to a proatlas (Castile et al., 1990). Castile and his group interpreted the defects such as the additional vertebra body, the rostrally extended notochord, and lateral ossification, as inductions anteriorly of more posterior characteristics (Castile et al., 1990).
Studies of homeobox function in developing *Xenopus laevis* embryos, involving either ectopic expression, blocking expression, or activity of genes, have offered some interesting insights into the function of HOM-C homeobox genes. The Xhox-1A homeobox gene, determined to be in the Antp class based on homology to *Drosophila* (Harvey et al., 1986, Regulski et al., 1985), appears to be involved in determining regional specification. Transcription constructs containing the coding region of this gene were used to produce mRNA for microinjection into fertilized embryos. Overabundance of this message in the developing embryo resulted in disruption of somitic mesoderm organization during development without loss of cell identity (Harvey and Melton, 1988).

The *Xenopus* homeobox containing gene Xhox3 seems to be involved in regional identity during development. This gene has been shown to be expressed as a gradient, with its highest level in the posterior region of the embryo (Ruiz i Altaba and Melton, 1989a). mRNA was produced and microinjected into the embryo in the anterior region of the embryo, where it is normally found at very low levels. The cells of the anterior region migrated during gastrulation to their proper destination, as determined by observation of movement of mesoderm cells from posterior to more anterior positions, but failed to differentiate into their normal anterior form. The head was deformed or absent, the central nervous system was deformed, the notochord was absent,
the somites fused along the middle of the embryo, and the axial length was reduced (Ruiz i Altaba and Melton, 1989b).

Absence of functional products of the Antp class *Xenopus* XlHbox1 gene, results in loss of segmentation, as well as homeotic like transformation. This gene codes for two proteins, one 82 amino acids longer at the N-terminal end, but both having the same homeobox and DNA binding specificity (Cho et al, 1988). Both proteins are expressed in the cervical region of the central nervous system, the neural crest and the mesoderm, with the expression of the shorter protein beginning more anteriorly in the mesoderm and the central nervous system, in comparison to the longer protein’s expression (Oliver et al., 1988). Microinjection of antibodies against a unique portion of the long protein resulted in the cervical spinal cord being transformed into a hindbrain like structure, i.e. into structures normally found just anterior to the cervical spinal cord. Over expression, or ectopic expression, of the long protein by injection of the mRNA coding for the long protein abolished segmentation. The injection of mRNA coding for the short protein resulted in abnormalities similar to the ones produced by injection of the antibodies to the long protein, except that the abnormalities extended more posteriorly. All three types of microinjection affected the development of the spinal nerves. They became a diffuse meshwork similar to cranial nerves instead of tight bundles characteristic of spinal nerves (Wright et al., 1989).
Organization of cells involved in embryogenesis of *Xenopus* can also be studied by implanting either treated or untreated embryonic tissue into the blastocoel cavity of an early gastrula embryo, followed by analysis of its affect. This was first determined by Spemann, who implanted dorsal lip cells into an Amphibian blastocoel and discovered that these cells would change the cell fate of the surrounding tissue, causing them to form highly organized axial structures like somites or the central nervous system (Spemann and Mangold, 1924). It has been shown that animal cap cells treated with a purified growth factor called *Xenopus* tissue culture mesoderm inducing factor (XTC-MIF) and implanted into the blastocoel of an early gastrula, would induce the formation of a secondary axis which included head structures (Cooke, 1989; Ruiz i Altaba and Melton, 1989).

A study of the *Xenopus* Abd-B homolog, XIHbox6, implicate this gene in specifying regional identity related to the activity of XTC-MIF. When animal caps cells treated with XTC-MIF were microinjected with mRNA coding for the XIHbox6 gene, not only were the normally anterodorsal inducing effects overridden upon implantation, but the host cells were recruited into producing secondary tails (Cho et al., 1991).

Recent analysis of XIHbox6 gene function suggests it has a role in cell, as well as regional identity. Niehrs and De Robertis microinjected mRNA coding for XIHbox6 into individual dorso-anterior blastomeres of 32 cell stage embryos, in order to examine the effect ectopic expression of the homeobox gene would
have on the fate of individual cells. They determined that injected cells that normally would give rise to brain failed to differentiate normally into cranial nerves. The tadpole cells derived from injected blastomeres gained head epidermal fate and lost anterior notochord fate. This appeared to be the result of altered cell migration during gastrulation, based on lineage analysis tracer using injected colloidal gold (Niehrs and DeRobertis, 1991).

Thus, all the studies discussed reinforce the idea that many genes with homeoboxes have similar functional roles in different organisms. Clusters that have been discovered are structurally related, and many of the genes are implicated in the determination of cell fate and regional specification, which contributes to segment pattern formation, and/or regional specification in the embryo during development.

An important variable in homeobox gene function is alternate transcription production, which may be regulated temporally and spatially. Alternate transcript production alone can give a homeobox gene different functions depending on when and where alternate transcripts are expressed. The Drosophila Abd-B gene, a 5' gene of the bithorax complex provides a good example. The Abd-B gene has been shown to be necessary for the correct morphogenesis of posterior abdominal segments A5-A9 (which corresponds to parasegments 10-15 of the embryo) of the adult fly. Abd-B is expressed in the posterior region of the embryo, consistent with its role in the morphogenesis of posterior abdominal segments (Harding et al., 1985; Casanova et al.).
Initial genetic analysis suggested that the Abd-B gene was bifunctional, and production of alternate transcripts contributed to its functional diversity. Two mutants were produced that complemented each other and mapped to the Abd-B gene. They were termed Class I or m mutants (for morphological), and Class II or r mutants (for regulatory). m mutant embryos exhibited A8 and A7 segments that looked more like the anterior segments A3-A6. r mutant embryos had defects in the adult terminalia (segment A9), in that the genitalia and analia are missing from both males and females. There was no recognizable effect on segments A5-A7 in these mutants (Casanova et al., 1986).

More genetic analysis confirmed that the r region regulated the expression of other genes, while the m gene had a more direct morphological affect. Complementation tests were carried out in animals with genetic backgrounds made specifically to emphasize the function of the two distinct regions of the Abd-B gene. Mutants missing the r region developed a denticle belt in their A9 region, which is an embryonic structure characteristic of more anterior segments. Also, r- mutants in a BXC background, having only a functional Ubx gene, had A9 segments resembling an A1 type that typically results from ectopic UBX expression. Finally, m+,r- mutants had A9 segments that had some abdominal segment characteristics that could be attributed to the ectopic expression of the m element in this region (Casanova et al., 1986).
The Abd-B gene has been characterized thoroughly by molecular analysis. Northern analysis with cDNA probes has detected up to 5 transcripts in a temporal series of embryonic mRNA samples, with sizes of approximately 7.8kb, 4.7kb, 4.3kb, 3.7kb, and 3.3kb (Zavortink and Sakonju, 1989; Kuziora and McGinnis, 1988; Celniker et al., 1989). Extensive screening of cloned genomic DNA restriction fragments with cDNA isolated from embryonic libraries has resulted in the mapping of a total of 8 exons to the Abd-B gene (Celniker et al., 1989). All 5 transcripts share the same 4 3' exons. These different messages can be categorized according to their unique 5' exons. Using this criteria, the transcripts can be put into three classes, with the 7.8kb message in one class (designated gamma class), the 3.7kb and 3.3kb transcript in another class (designated beta class) and the 4.7kb and 4.3 kb messages making up the final class (designated alpha class). The presence of two different messages within the gamma and beta groups is probably due to the use of an alternate poly-A site. The longer messages of the two classes hybridized to a genomic probe made from sequence 3' of the poly-A signals for the shorter messages (Kuziora and McGinnis, 1988).

Determination of the localization of expression of the different mRNA classes, in addition to the thorough genetic analysis, has helped classify the transcripts as relative to the m(morphological) or r(regulatory) genetic determinants. Probes specific for each class, as confirmed by Northern analysis, were used for in situ hybridizations that showed that all three classes were expressed in...
ectoderm and mesoderm. The alpha class, containing the 4.7kb and 4.3kb messages, were localized to parasegments 11 to 15. The 3.7kb and 3.3kb messages of the beta class were localized to the parasegments 14 and 15. The expression of the gamma class was limited to parasegment 15 (Kuziora and McGinnis, 1988). Kuziora and McGinnis propose that one or both of the alpha class transcripts confer the morphological function of the Abd-B gene, based on their expression pattern and the fact that m-r+ mutants map to the genomic restriction fragment encoding the exon unique to this class. They also suggest that the beta and or the gamma transcripts confer the regulatory function of the gene, based on their profile of spatial expression.

Other research supports the above conclusions. Zavortink and Sakonju used 5' ends of cDNA's representative of alpha and beta RNA to do primer extensions, using S1 nuclease protection data obtained from genomic templates to corroborate their results. They mapped different initiation sites for the different RNA's, showing that they are transcribed from different promoters. They also determined that the 411 base pair deletion of the m-r+ mutant removes the transcription initiation site of the alpha class 4.6kb RNA. This initiation site is downstream of the beta/gamma site, and is located within an intron of the other two RNA classes (Zavortink and Sakonju, 1989). This indicates that the m-r+ deletion most likely interrupts the expression of the alpha messages, while the other messages can still being expressed. This
supports the idea that alpha class transcripts are responsible for the morphological function of the Abd-B gene.

Other work supports the contention that the gamma Abd-B product, whose expression is confined to parasegment 14, is responsible for Abd-B’s regulatory function (DeLorenzi et al., 1988). First they determined which regions of the BX-C complex were transcriptionally active during embryogenesis by carrying out a "reverse Northern" analysis. Restriction fragments of clones from this 5' region were hybridized to labeled RNA produced during embryogenesis. Four active regions were identified, and the two which mapped closest to r mutant deletions were selected to screen a cDNA library constructed from embryonic RNA. They isolated a clone that contained the Abd-B homeobox, as well as all of the putative upstream sequence and used this to construct a probe specific for this RNA. In situ analysis revealed that this RNA's expression was confined to parasegment 14, which is the area affected by the Abd-B gene r mutants. In addition to the above data, three of the m+r- mutants that have been characterized have rearrangement mutations which would be disruptive to the transcription of the RNA discussed here (DeLorenzi et al., 1988).

It seems apparent that alternate transcript production allows the Abd-B gene to carry out multiple functions. Is this type of regulation evolutionarily conserved, or is it a new function among HOM-C homeobox genes, whose sequence homology suggests they are descendants of an ancestral Abd-B like gene? Because each of the 4 clusters of HOM-C homeobox genes found in
the genome of mouse and man contain an Abd-B subfamily gene, based on physical alignment and sequence homology, there are a number of vertebrate Abd-B class genes to examine. The murine genes considered here are termed Hox-1.7, Hox-2.5, Hox-3.2, and Hox-5.2. Accumulating research discussed below suggests the vertebrate Abd-B homologs utilize some of the same molecular mechanisms as the Drosophila Abd-B gene and this may be functionally significant.

Northern of embryonic and adult mRNA have detected alternate transcripts produced from Hox-1.7, Hox-2.5 and Hox-3.2 (Rubin et al., 1987; Erselius et al., 1990; Bogarad et al., 1989). As with the Drosophila Abd-B gene, the embryonic expression of these genes is restricted to more posterior regions of the embryo in a colinear fashion relative to other genes of the complexes. Northern analysis has not been reported for Hox-5.2, but spatial localization of the transcripts of all four genes have been characterized to some degree.

Localization of expression for the Hox-1.7 gene has been examined by temporal series and morphological dissection. Northern blot analysis was carried out on mRNA from from 11.5 to 16.5 day old embryos, as well as from some adult tissues. In addition, 12.5 day embryos were dissected into brain, anterior spinal cord, posterior spinal cord, and carcass, and their respective mRNA's were isolated and analyzed. The two largest messages, 3.9kb and 2.5kb, were detected, with the 2.5kb messages being most abundant, in 11.5 and 12.5 day embryos. Both messages decreased thereafter and were absent
by day 16.5. Three transcripts, including the two discussed above and, a third message of 1.9kb, were found in adult tissues of the kidney, spinal cord, heart, and spleen, but were absent in the brain or uterus. All three transcripts were detected in the embryonic posterior spinal cord, with the 2.5kb message being the most abundant. Very minute amounts of all messages were detected in the embryonic anterior spinal cord.

No Hox-1.7 mRNA was detected in the embryonic brain or carcass (Rubin et al., 1987). Hox-1.7 mRNA appeared to be primarily restricted to posterior spinal cord in 12.5 day embryos, indicative of the posterior localization that is expected for products of Abd-B homologues.

Products of the Hox-2.5 and Hox-3.2 Abd-B homologues have been examined using in situ hybridization with 35-S labeled RNA as well as Northern blot analysis, to determine the temporal expression pattern and the spatial localization of expression. mRNA from different embryonic time points, as well as from various adult tissues, was isolated for Northern blot analysis. The larger 3.0kb transcript of the Hox-2.5 gene was first detected in 8.5 day embryos, was found to increase by day 10.5 and persisted through the 13.5 day. A smaller 2.3kb transcript appeared by day 13.5. Both transcripts were also detected in the kidney and the spinal cord, but not in the brain, liver, lung, heart, ovary, or testis. In situ analysis on sections from various embryonic stages revealed changing regionalized expression in tissues of the spinal cord and mesoderm. The expression of the Hox-2.5 gene was restricted to regions
which are most posterior of the expression of any of the genes in the Hox-2 cluster (Bogarad et al., 1989).

For Hox-3.2, the 3.2kb, 1.9kb, and 1.5kb transcripts of the gene were detected from day 9 through day 15 of embryogenesis, with the smallest transcripts being in the highest abundance. In adult tissues, Hox-3.2 transcripts were found only in the kidney, where the two smaller transcripts were also detected. This gene also displayed the most posterior expression of any of the genes in the Hox-3 cluster. It was expressed posterior to the third thoracic prevertebrate in the central nervous system, and posterior to the ninth thoracic segment in sclerotomes (Erselius et al., 1990).

Extensive in situ analysis has been carried out on the Hox-5.2 gene. These results have shown that the Hox-5.2 gene is expressed in the posterior regions of the spinal cord, and in sclerotomes posterior to pre-vertebrate 21 and 22 (Duboule and Dolle., 1989).

Molecular analysis suggests that, like the Drosophila Abd-B gene, alternate splicing is responsible for different size transcripts in many of the vertebrate Abd-B homologues. Mapping of isolated cDNA's to their respective genomic regions has been carried for three of the genes. A 0.8kb Hox-1.7 cDNA sequence is not continuously represented on the genomic DNA, indicating the presence of a splice site (Rubin et al., 1987). More recently, two cDNA clones representing the Hox-1.7 homologue in the guinea pig have been isolated. One may represent a complete region of an mRNA, since it has an initiation codon.
that is in frame with the homeobox. Both guinea pig clones are identical in their 3' region, but their sequences diverge significantly 12 codons 5' of the homeobox. This, along with Southern blot analysis, indicate that there is an alternate splice site just 5' of the homeobox which results in the production of different transcripts depending on which 5' exon is utilized (Rubin et al., 1990).

The Hox-2.5 gene has a similar structure. An isolated cDNA that contains the homeobox with a 5' open reading frame, diverges from the sequence of its genomic coding region 12 codons 5' of the homeobox at a splice acceptor site and the coding sequence for the cDNA appears 2.9kb upstream at a splice donor site. A probe derived from sequence just upstream of the donor site recognizes only the longer of two transcripts produced by this gene, indicating the two different size transcripts are produced by alternate splicing (Bogarad et al., 1989).

The Hox-3.2 gene produces three different transcripts of sizes 3.2kb, 1.9kb, and 1.4kb. Alignment of a 1.4kb cDNA to the corresponding genomic region reveals a 1.7kb intron that starts 12 codons upstream from the homeobox. A probe made from the 5' part of the cDNA encoded by part of the upstream genomic sequence hybridizes only to the 1.4kb and 1.9kb transcripts, indicating that longer 3.2kb transcript does not contain this exon, but is produced by alternate splicing.

The Drosophila Abd-B gene and at least one of the Hox Abd-B homologs also utilize alternate poly-A signals. For the Drosophila gene, a 400 base pair
difference in the 3' region of the mRNA's within the alpha and gamma class can be attributed to the use of alternate poly-A signals (Kuziora and McGinnis, 1988). It appears that two Hox-3.2 transcripts use alternate poly-A signals. Two different cDNA clones representing the Hox 3.2 gene, utilize different polyadenylation signals 450bp apart. This indicates that the size difference in the 1.5kb and 1.9kb transcripts maybe due to poly-A site variation (Erselius et al., 1990). Closer analysis of the other Hox Abd-B homologues may uncover similar mechanisms.

There seem to be some similarities between the Abd-B gene and the Murine Hox homeobox genes that are probably functionally significant. The expression of all the Hox genes and the Drosophila Abd-B gene is restricted to more posterior regions of the embryo. The genes all exhibit colinearity, with the 5' genes of the respective clusters being expressed in more posterior regions than the 3' genes. This type of colinearity is also exhibited by Hox-5.2 and the whole Hox-5 cluster during limb pattern formation (Dolle et al., 1989). The Abd-B class genes of the vertebrates and the Abd-B gene of Drosophila all produce multiple transcripts in a temporally regulated fashion, including production of different mRNA combinations and abundances during development, and in adult tissues. All the mechanisms discussed above could contribute to multifunctionality of the Hox Abd-B homologue genes, in a way similar to the Drosophila Abd-B gene.
The Abd-B sea urchin homologue TgHBox4 produces two different transcripts of 4.4kb and 3.7kb during development, and a third transcript of 3.5kb in some adult tissues. Analysis of a genomic clone has revealed a homeobox sequence that has been shown to have 90%, 88%, 88%, and 87% amino acid identity to the Hox-1.7, 2.5, 3.2 and 5.2 homeoboxes, respectively (Dolecki et al., 1988; Vansant, unpublished results). TgHBox4 shows 73% amino acid identity to the Drosophila Abd-B homeobox (Dolecki et al., 1988). High homology between the homeoboxes of the Hox Abd-B homologs, the Drosophila Abd-B gene, and the sea urchin Abd-B class gene are likely due to functional similarities that depend on the conservation of certain sequences and mechanisms. It suggests that much of the sequence of the homeobox of the Abd-B genes is important for the production of functional proteins. Molecular similarities may be of functional significance also. The mechanisms of function of the Abd-B homolog in the sea urchin is of particular interest in terms of comparison to the same homologs function in the vertebrates. The echinoderms and the vertebrates are more evolutionarily related than the echinoderms and the arthropods. The sea urchin evolved after the separation of the deuterostome and protostome evolutionary lines. Sea urchin embryos are one of the most, if not the most simple life form in the evolutionary line from ancestral deuterostome to vertebrates. Functional correlations between urchin homeobox genes and vertebrate homeobox genes should therefore be more extensive than correlations between Drosophila and vertebrate homeobox gene
function. This statement will seem even more reasonable, if homeobox genes in sea urchins are found to have been recruited in fate specifications related to the bilaterally symmetric body plan of the developing embryo. I would like to carry out a more detailed analysis of the TgHBox4 gene. I plan to determine its expression pattern in the developing embryo and determine the mechanism of alternate transcript production. By analyzing this gene and comparing the results with those obtained in the study of Abd-B genes I will propose the function or functions this gene may have in Tripnuestes gratilla. This study should at least illuminate the importance of conserved features found among Abd-B like genes.
CHAPTER II

MATERIALS AND METHODS

DNA gels

DNA was quantitated and sized on gels ranging from .7 to 1.5% agarose as follows. First the appropriate amount of SeaKem agarose (FMC Corporation) was dissolved in Tris-Acetate buffer (TAE)(0.04M Tris-Acetate 0.002M EDTA pH 8.0) by heating the solution until boiling. 20ml solutions were made for mini-gels (Model aSH quick screening electrophoresis, IBI INC.), and 100ml was used for standard size gels (Model MPH multi-purpose gel electrophoresis, IBI INC.). These solutions were poured into horizontal gel casting stands at room temperature with the sample wells being formed with teflon gel combs. The samples were prepared as follows: The appropriate amount of 10X gel loading buffer (10mM Tris, 1mM EDTA, 50% glycerol, 0.1% bromphenol blue) was added to DNA samples that were dissolved in ddH2O or 10mM TRIS 1mM EDTA pH 7.5 (TE). These samples were heated at 69 degrees C. for 10 minutes and quick chilled on ice for 5 minutes before loading into the sample wells. The gels were run in TAE buffer at 100V at room temperature for 50 minutes. After running the gels they were stained with ethidium bromide at a concentration of 50ug/ml in ddH2O for 10 minutes, rinsed with ddH2O, and then photographed under ultraviolet illumination. The photograph was taken with the
545 Polaroid Land camera. The film used was either type 55 or type 57 Polaroid Land film.

**Plasmid minipreps**

1.5-2.0ml of an overnight culture of *E. coli* containing the plasmid to be purified was transferred to an eppendorf tube and centrifuged at 17,000R.P.M.’s in an eppendorf centrifuge (Brinkman Model 5414). All the supernatant was removed after centrifugation and the cell pellet was resuspended in 100ul glucose buffer (50mM glucose, 10mM EDTA-tetrasodium salt, 25mM TRIS, pH 8.0, 2mg/ml lysozyme) and left at 25 degrees Celsius for 5 minutes. 200ul of 1% Sodium Dodecyl Sulfate (SDS) and 0.1N NaOH were then added and carefully combined with the resuspended cells which were then put on ice for 5 minutes. Next 150ul of 3M Na/5M Acetate were added to the cells, which were carefully mixed and put back on ice for 5 minutes. The solution was then spun at 4 degrees Celsius at 17,000R.P.M.’s for 5 minutes in an eppendorf centrifuge (extraction). After spinning the supernatant was removed and placed in a clean eppendorf tube to which was added 1/2 volumes of phenol and chloroform. The mix was then vortexed vigorously and then spun for 5 minutes at 17,000R.P.M.’s in an eppendorf centrifuge. After spinning, the aqueous top layer was removed and transferred to a clean eppendorf tube. Two volumes of 95% ethanol were added to the tubes and they were placed at -20 degrees Celsius from 20 minutes to overnight. After the -20 incubation the tube was
spun at 4 degrees Celcius at 17,000R.P.M.'s for 10 minutes in an eppendorf centrifuge. The ethanol was then poured out and replaced with 200ul of 70% ethanol. The tube was vortexed for 5 seconds and then spun at 4 degrees Celcius at 17,000R.P.M.'s for 5 minutes. The ethanol was removed from the tubes using a Pipetteman (eppendorf), being careful as to not disturb the DNA pellet. The DNA pellet was dried by placing the eppendorf tube in a SpeedVac concentrator (Savant) under constant vacuum with heat for 20 minutes. After drying, the DNA pellet was resuspended in 20ul of T.E. with RNase at a concentration of 50ug/ml and incubated at 37 degrees Celcius for 15 minutes. The concentration and integrity of the sample was checked by running 1/80 of the preparation on an agarose gel as previously described.

DNA Precipitations

First the DNA solution was phenol:sevag (24 parts chloroform, 1 part isoamylalcohol) extracted with equal volumes of each. Then the aqueous phase was transferred to a new microfuge tube and brought to a concentration of 0.3M NaAcetate and a pH of 5. 2 volumes of ice cold 95% ethanol was then added to the solution, which was placed at -20 degrees Celcius for 30 minutes to overnight. The solution was then spun in a microfuge for 10 minutes at 4 degrees Celcius. The supernatant was carefully removed and the DNA pellet rinsed with 200ul of ice cold 70% ethanol. The tube was spun for five minutes in a microfuge at 4 degrees Celcius. The ethanol was removed and the DNA
pellet dried in a SpeedVac concentrator (Savant) by spinning the tube for 20 minutes with heat and under a constant vacuum. The DNA was then redissolved in the appropriate buffer.

**Restriction Digests**

Typically 1-5ug of DNA were digested in a volume ranging from 10ul to 50ul. The composition of the buffer used for each restriction digest was obtained using one of the four standard 10X buffers supplied by New England Biolabs (NEB). In some cases a NEB buffer custom made for a specific enzyme was used. The NEB buffer used was determined by the recommendations found in the NEB catalog for each enzyme. In addition to the reagents supplied by the 10X NEB buffers, spermidine was added to a final concentration of 1mM and BSA was added to a final concentration of 100ug/ul in each reaction. Double digest were handled two ways. If the buffer conditions for both enzymes was compatible 1 enzyme was reacted for 1 hour at the specified incubation temperature. This enzymes was then heated inactivated at 69 degrees if possible. The second enzyme was then added and incubated for 1 hour to overnight at the specified incubation temperature. If buffer conditions were incompatible for a double digest, one enzyme reaction was carried out, the DNA precipitated as described above, and resuspended in the buffered solution appropriate for the second enzyme reaction. The second enzyme
would be added. After this, the second reaction would be allowed to proceed as described above.

Ligations

Digests were loaded in agarose gels and electrophoresed as previously described with the following modifications. Gels were prepared using SeaPlaque low gelling temperature agarose and were poured at 4 degrees Celcius. These gels were ran at 75 Volts for at least an hour at 4 degrees Celcius. After ethidium bromide staining and rinsing with ddH2O, the appropriate bands were removed with a single edge razor. Gel slices were heated to 65 degrees C. to dissolve them so their volume could be determined, and then placed at 37 degrees C. to cool before being used. A known mass of DNA was loaded on the gel. The concentration of the DNA in the agarose isolate was determined by dividing the percentage of the mass of DNA loaded that represented the restriction fragment isolate by the volume of the isolated gel slice. Unless stated otherwise, ligation reactions were carried out at a ratio of 3 moles of insert to 1 mole of vector in 660mM Tris-HCl pH 7.6, 100mM MgCl₂, 10mM DTT, 2mg/ml gelatin, 1mM ATP, and 2 units T-4 DNA ligase in a volume range of 20 to 50μl at 14 degrees C. overnight. The next day an E. coli strain of choice was transformed with 1/2 of the ligation mix (Mandel and Higa, 1970). If a low backround of non-insert containing vectors was expected, the transformants were screened using plates with antibiotics.
resistant to if they contained the vector. If a high background of insert minus vector containing bacteria was expected, the antibiotic plates were supplemented with X-galactosidase (X-gal) and IPTG at concentrations of 2% and 0.1M, respectively. Bacteria containing vector with inserts would then appear as white colonies and bacteria with just vector would appear as blue colonies. Alternatively, bacteria were sometimes screened for the desired clone using polymerase chain reaction (PCR) amplification as described below.

**Probe Production**

Random prime labeling

This labeling technique was used for DNA fragments that were 400 base pairs or larger. First the appropriate digest was carried out as previously described, to produce the desired DNA fragment. The DNA fragment for labeling was isolated, and its DNA concentration determined as previously described in the ligation section. A volume of the agarose containing from 5 to 50ng of DNA was diluted at a ratio of 3ml of autoclaved double distilled water per gram of gel. The DNA was denatured by boiling the above solution for 7 minutes. The solution was placed in a 37 degree Celsius water bath. Labeling was carried out using the United States Biochemical (USB) Random Primed DNA Labeling Kit. 5ul of reaction mixture (includes random hexanucleotides), 7.5ul of dNTP mix(1:1:1 mixture of each dNTP), denatured DNA solution, and autoclaved distilled water totalling 31.5ul, and 5ul of alpha-32 phosphate dCTP
(3000Ci/mmol) were combined. 1ul of Klenow enzyme at a concentration of 2
Units/ul was added, and the mix incubated at 37 degrees Celcius for at 30
minutes to 5 hours. The reaction was terminated by heating the mix for 10
minutes at 65 degrees Celcius. The labeled probe produced was isolated with
a 1ml spin column loaded and packed with G-50-80 sephadex (Sigma) that was
equilibrated with 10mM Tris-HCl, pH 7.6, 1mM EDTA, 100mM NaCl (TEN or
STE). The labeling reaction was diluted to 100ul, loaded on the spin column
and the column was spun at 2000R.P.M.'s for 2 minutes. The labeled probe
was collected in a centrifuge tube and 2ul added to 2ml of Ecolume (ICN
Biomedicals). The radioactivity of the probe sample in Ecolume was measured
with a LS 7000 Scintillation counter (Beckman).

End labeling
This labeling reaction was used for synthetically produced oligonucleotides
with sizes ranging from 25 to 35 base pairs. 2ul of 500mM Tris-HCl, pH 7.6,
100mM MgCl2, 1mg/ml (10X buffer), 1ul of 200mM dithiothreitol (DTT)
(FisherBiotech), 1.5ul oligonucleotide(100ng/ul), 5ul of gamma-32-phosphate
(10uCi/ul), 2ul 1mM spermidine, .5ul of T4 polynucleotide kinase (10U/ul)(NEB),
and 6ul ddH2O were combined and incubated at 37 degrees Celcius for 30
minutes. After 30 minutes an additional 5 Units of T4 polynucleotide kinase was
added and the reaction was allowed to proceed for another 30 minutes at 37
degrees Celcius. The reaction was inactivated by heating the solution at 68
degrees Celcius for 5 minutes. The reaction mix was brought to a volume of

47
100ul with ddH2O. The labeled probe was then isolated and the radioactivity of a sample was determined as previously described in the random prime labeling section.

**PCR Screening**

Template source

**For Screening of Bacterial Transformants**

Bacterial colonies were picked from the transformation plates and grown overnight at 37 degrees Celsius in LB media containing the appropriate antibiotic. 1ul of a turbid culture was removed and used as a source of DNA template.

**For Screening of Lambda clones**

A Lambda clone DNA sample was diluted to a concentration of 1ng/ul and 1ul was removed and used as a source of DNA template.

Amplification

DNA was amplified from 1ul of DNA containing solution in 50ul of PCR buffer (50mM KCl, 10mM Tris-HCl, pH 8.8 at 25 degrees Celsius, 1.5mM MgCl2, and 0.1% Triton X-100). The reaction included 0.2mM of each of the dNTP’s, the appropriate primers (vector primers flanking insert for bacterial screening and one lambda specific primer and one insert specific primer for lambda clone screening) at a final concentration of .5pmol/ul, and TAQ polymerase (Promega) at a final concentration of 0.05Units/ul. The reactions were layered
with mineral oil to prevent evaporation of samples during the reaction. The PCR reaction was carried out in a DNA Thermal Cycler (Perkin Elmer Cetus) using a three segment program that was repeated 25 to 30 times. A typical program was as follows: a first segment that was a 94 degree Celsius incubation for 1 minute; a second segment that was a 50 degree Celsius incubation for 1 minute and 20 seconds; and a third segment that was a 72 degree Celsius incubation for 2 minutes. To visualize the products, 1/10 of each reaction was run out on an agarose gel as previously described.

cDNA and genomic DNA library screenings

Phage plates

Three libraries were screened. Two different cDNA libraries and one genomic DNA library. One cDNA library was constructed by Invitrogen using the lambda gt11 vector and poly A+ RNA from blastula stage *Tripneustes gratilla* embryos. The other cDNA library was constructed by Marcus Knoll using the lambda ZAP-cDNA synthesis vector kit and poly A+ RNA from the large intestine of adult *Tripneustes gratilla* urchins. The genomic library was constructed by Stratagene using the lambda gt11 vector and a urchin sperm DNA that had been partially digested with EcoR1. 5X10^4 phage were plated onto 150X15mm plastic petri dishes (Fisher).
lambda gt11 cDNA library plating

First a volume of a library aliquot was incubated with 100ul of an overnight culture of *E. coli* strain Y1088 grown in LB for 20 minutes at 37 degrees C. This was then mixed with 5ml of prewarmed (42 degrees C.) LB-top agarose and plated out on prewarmed (37 degrees C.) 150mm LB plates. After hardening at room temperature these were incubated overnight at 37 degrees C..

lambda ZAP cDNA library plating

First a volume of a library aliquot representing 50,000 plaque forming units (P.F.U.'s) was incubated with 600ul of .5 O.D.600 *E. coli* strain PLK-F', that had been grown in NZY broth with .2% maltose for 15 minutes at 37 degrees C.. This was then mixed with 5ml of prewarmed (42 degrees C.) NZY top agarose and plated out on prewarmed (37 degrees C.) 150mm NZY plates. After hardening at room temperature these plates were incubated for 8 to 10 hours at 37 degrees C..

lambda gt11 genomic DNA library plating

First a volume of a library aliquot was incubated with 600UL of .5 O.D.600 *E. coli* Y1090 strain that had been grown in NZY broth that included 0.2% maltose and 10mM MgSO4 for 15 minutes at 37 degrees Celcius. This was then mixed with 6ml of prewarmed (42 degrees C.) NZY top agarose and plated out on prewarmed (37 degree Celcius) 150mm NZY plates. The agarose was allowed to harden at room temperature. After hardening, the plates were incubated for 8 to 10 hours at 37 degrees Celcius.
Transfer to nitrocellulose

The phage plates were chilled at 4 degrees C. for at least one hour before the lifts were carried out. Duplicate lifts were made using 0.45μm nitrocellulose 150X15mm circular filters (Micron Separations Inc.) with the first filter being left on the plate for 3 minutes and the second filter remaining for 5 minutes. Three asymmetric holes were put in the filters and into the agar plates with an 18 gauge needle so that the filters could be oriented with the plates to identify regions with positive signals (see Hybridization below). The DNA on the filters was then denatured by exposure to a solution of 0.1N NaOH and 1.5M NaCl for at least 30 seconds. The filters were then neutralized in 10X SET (0.3M tris-HCl pH 7.5, 1.5 M NaCl, 20mM EDTA) for at least one minute. Finally the filters were washed with two 4X set rinses for at least one minute each. The filters were then allowed to dry for at least 30 minutes at room temperature, plaque side up on Whatman #1 filter paper. The filters were then turned over labeled with a #2 pencil, stacked between glassine sheets, wrapped in Whatman #1 paper, wrapped in aluminum foil and then baked in a vacuum oven for 1.5 hours at 80 degrees C. After baking, the filters were either stored as is at 4 degrees C. or used immediately in hybridization reactions.
Hybridization reactions

Hybridization reactions were carried out in a Robbins scientific hybridization oven with rotating borosilicate GL 45 Schott tubes. First the lifts were rinsed in 6X SET for five minutes at room temperature in a pyrex baking dish. The filters were then pre-hybridized in the oven at 68 degrees C. in a 20ml solution consisting of 6X SET, 10X Denhardt’s, 0.1% SDS and 0.1% NaPPi for at least 1 hour. This solution was removed from the tubes and replaced with a second pre-hybridization solution of 10ml consisting of 6X SET, 5X Denhardt’s, 0.1% SDS, 500ug of yeast torula tRNA that had been heat denatured at 70 degrees C and 100ug of E. coli DNA that had been denatured with boiling for 10 minutes. The lifts were prehybridized in this solution for at 68 degrees C. for at least 1 hour. After pre-hybridization, the appropriate ³²-P labeled probe was added to the second prehybridization solution at a concentration of 10⁶ counts per minute per ml. of solution. The hybridization was carried out at 68 degrees C. for at least 8 hours. After hybridization, the lifts were rinsed in solutions of decreasing salt concentrations with SDS. First two 6X SET, 0.1% SDS washes were carried out at room temperature for 5 minutes followed by one wash with the same solution at 68 degrees C. for 10 minutes. These were followed by successive 10 minute washes at 68 degrees C. with 4.5X SET, 0.075% SDS; 2.0X SET, 0.033% SDS; and 1.5X SET, 0.025% SDS. If the blot was determined to be too radioactive after this series (probable low signal to background ratio), the final wash was repeated until acceptable radioactive levels were achieved.
After washing, the lifts were mounted five at a time on #1 Whatman filter paper and wrapped with saran wrap. These were then put against X-ray film in a film cassette holder with intensifying screens and put at -70 degrees C. for at least 12 hours.

Purification of positive plaques

Areas on the plaque plates containing a hybridization positive clone were identified by aligning exposed areas on the autoradiograph to these areas. Once identified these areas would be removed from the plate using the large end of a sterile Pasteur pipette and placed in 1ml of SM containing 50ul of chloroform. Next phage were plated in a manner similar to the procedures previously described. The eluted phage were diluted to $10^4$ of the stock concentration, and a 1ul, 10ul, and 50ul aliquot of this dilution were incubated with the appropriate amount and strain of *E. coli* cells, mixed with 3ml of top agarose, plated on 15X100mm plastic petri dishes (Fisher) of the appropriate type and incubated overnight at 37 degrees C. These plates were then chilled and the plaques transferred to nitrocellulose, and subsequently reacted to the appropriate radioactive probe(s) as described above. Based on an autoradiograph to the posthybridized lifts, a well isolated plaque was picked using the small end of a sterile Pasteur pipette tip. This isolate was placed in 1ml of SM buffer with 50ul of chloroform and stored at 4 degrees C.
Positive clone DNA isolation and characterization

lambda ZAP clones

Clones determined to be positive that were obtained from the library constructed with the lambda ZAP-cDNA synthesis kit, were characterized according to the protocol supplied by Stratagene, with certain modifications recommended by Marcus Noll. 100μl of the SM solution containing the positive phage particles (>1X10^5 phage particles) was combined with 100μl of an overnight culture of XL-1-Blue cells (Stratagene), 100μl of a MgCa solution, and 1μl of R408 helper phage (>7.5X10^10 pfu/ml) (Stratagene) in a 12ml culture tube. After a 10 minute incubation period at 37 degrees Celsius, 1ml of NYZ media was added and the mixture was incubated for an additional 2 hours at 37 degrees Celsius with shaking, using the Lab-Line incubator shaker (Labline instruments). The tube was then heated at 70 degrees Celsius for 20 minutes and then spun for 5 minutes at 4000g in the RC-5 Superspeed Refrigerated Centrifuge using the HB-4 rotor (Sorvall Centrifuges-Dupont). The supernatant was then decanted into a sterile culture tube. This stock containing the pBluescript phagemid packaged as filamentous phage particle was then used to infect E. coli cells. 200μl of this phage stock was added to 200μl of O.D.=1.0 XL1-Blue cells and 1μl of 1X10^3 dilution of this phage stock was added to another 200μl of XL1-Blue cells and both solutions were incubated at 37 degrees Celsius for 15 minutes. 1μl, 5μl, and 50μl of each of the infected set of cells was plated out onto LB\100ug\ul Ampicillin plates and incubated,
inverted, overnight at 37 degrees Celsius. An isolated colony was picked from one of the plates and grown overnight in 10ml of LB broth with Ampicillin at a concentration of 100ug/\text{ul}. The phagemid was retrieved from the cells using the plasmid mini-prep method as previously described, run out on a 1% agarose gel, and stained as previously described to determine the plasmids approximate concentration, using 500ng of lambda DNA digested with HindIII and run on the gel as a reference. 500ng of plasmid was then restriction enzyme digested as previously described using the enzymes EcoRI and XhoI to release the cDNA insert. This digest was then run out on a 1% agarose gel to determine the size of the insert again using 500ng of lambda DNA digested with HindIII as a reference.

lambda gt11 clones

The DNA was isolated from these clones using the plate lysate method. An agarose plug including a well isolated positive plaque was eluted in 1ml of SM for two hours with gentle mixing. 100ul of the eluate was used to infect 600ul of an overnight culture of Y1090. The cell eluate mixture was placed at 37 degrees Celsius for 15 minutes, mixed with 6ml of SEA-KEM agarose (FMC) that had been preheated to 42 degrees Celsius and spread out on 150mm plates that were prewarmed to 37 degrees Celsius. After the agarose solidified, the plates were incubated at 37 degrees Celsius for 8-10 hours. The plaque covered plates were then layered with 12ml of SM and placed on a rotary shaker for 2 hours to elute the phage. The SM was then transferred to 12ml
snap cap tubes and spun at 9000 R.P.M.'s in a Sorval HB-4 rotor at 4 degrees Celsius for ten minutes. The supernatants were transferred to new tubes and RNase A and DNase I were both added to a final concentration of 5ug/ml. The tubes were then placed at 37 degrees Celsius for 30 minutes. Next, an equal volume of 20% PEG (polyethylene glycol) 8000, 2M NaCl in SM was added to the supernatants. These were then incubated in ice in a 4 degree Celsius cold room for at least 1 hour. The samples were then spun at 9500 R.P.M's for 20 minutes in the Sorval HB-4 rotor at 4 degrees Celsius. The supernatants were poured off and the tubes inverted at room temperature for 10 minutes. The pellets were resuspended in 500ul of SM and dissolved, by heating at 37 degrees Celsius with slight agitation, and placed in microfuge tubes. The solutions were then extracted with an equal volume of phenol by vortexing to bring the PEG to the interfaces. After spinning the tubes for 3 minutes in the microcentrifuge, the aqueous layers were removed and transferred to fresh 1.5ml microfuge tubes. 14ul of 0.5M EDTA pH 8.0 was added to each tube and then they were briefly vortexed. 3ul of 25% SDS was added, the tubes vortexed and then quickly placed at 68 degrees Celsius to incubate for 15 minutes. The solutions were extracted with phenol, with sevag added to the phenol, and then with just sevag. 1/20 of the volume of each solution of 3M NaAc was added to each tube and the DNA was precipitated by adding an equal volume of ice cold isopropanol to each tube and placing the tubes at -20 degrees Celsius overnight. The DNA was spun down for 20 minutes at 4
degrees Celsius in a microcentrifuge and then rinsed with 70% isopropanol. Finally, the DNA was dried and dissolved in T.E. with RNase at a concentration of 40ug/ml.

Southern Blots

DNA to be transferred to a solid support was first run out on an agarose gel as previously described. Afterwards, the gel was soaked in a denaturing solution composed of 0.5M NaOH and 1.5M NaCl for 45 to 60 minutes. Next the gel was transferred to a neutralization solution made up of 1.0M Tris-HCl pH 5.0, 3.0M NaCl and soaked for 45 minutes. After this the gel was placed on top of a paper wick (Whatmann #1) that was placed perpendicular on a 10cm wide glass plate. The glass plate was placed across a 20cm*20cm*5cm pyrex dish, with the ends of the paper wick touching the bottom of the pyrex dish. The pyrex dish was filled halfway with 20X SSC. Next, a piece of 0.45um Nitrocellulose was cut to the size of the gel, pre-wetted with ddH2O, and placed on the gel. The Nitrocellulose filter was covered with a piece of Whatman #1 paper that had been saturated with 2X SSC. Air bubbles between the gel and the filter were removed by gently rolling a sterile 15ml disposable test tube (Fisherbrand) over the Whatman #1 paper covering the Nitrocellulose. Next paper towels cut to fit the size of the gel were stacked onto the Whatman paper to a height of approximately 10cm. This was covered with a 15cm*15cm glass plate that was weighted down with a fluid filled flask whose mass did not
exceed 500 grams. The gel was surrounded by previously exposed autoradiographic film to insure capillary action was exclusively through the gel and filter. This was allowed to stand overnight. The next day the towels and the Whatman paper were removed from the Nitrocellulose, and the location of the sample wells marked with a #2 pencil. The Nitrocellulose was then removed, rinsed in 2X SSC, and then baked in a vacuum oven at 80 degrees Celsius for 2.0 hours. The blot was then stored at 4 degrees Celsius until it was needed for hybridization. Hybridization was carried out as previously described for plaque containing Nitrocellulose lifts.

**DNA Sequencing**

DNA restriction fragments were subcloned into pBluescript II SK+(Stratagene). In some cases these templates were modified for internal sequencing by the creation of a nested set of deletions using exonuclease III (Ausubel et al., 1992). Screening of nested deletion clones was expedited using the PCR technique previously described. Templates were sequenced by the modified dideoxy chain termination method using the version 2.0 sequenase DNA sequencing kit supplied by UNITED STATES BIOCHEMICALS (Tabor and Richardson, 1987). Junctions of expression constructs and DNA regions that were unobtainable by insert flanking primer reactions were sequenced with the aid of synthetic primers supplied by the University of Hawaii Biotechnology-Molecular Biology Instrumentation Facilities. Sequencing reactions were
electrophoresed on 7M urea, 6% polyacrylamide gels (Biggen et al., 1983; Sheen and Seed, 1988; Parkison and Cheng, 1989), and autoradiographed. Sequence obtained was analyzed using the programs of the Intelligenetics PCGene software package.

Protein gels using Mini-PROTEAN 11 Dual Slab Cell

Minigels were prepared using casting stands, glass plates, spacers, and well combs that were supplied by the manufacturer (Biorad). Gels solutions were made up using the following stock solutions: a) stock acrylamide solution; 73 grams of 2x purified acrylamide crystals were dissolved in 250 ml of ddH2O along with 2 grams of bis-acrylamide to make a stock solution that was 30% w/v acrylamide with a crosslinking concentration of 2.7% w/w. b) stock separating gel buffer; 1.0 gram of SDS was dissolved in ddH2O along with 45.5 grams of TRIZMA-base and brought to a pH of 8.8 with HCl before bringing the final volume to 250 ml. c) stock stacking gel buffer; 1.0 gram of SDS was dissolved in ddH2O along with 15.1 grams of TRIZMA-base and brought to a pH of 6.8 with HCl before bringing the final volume to 250 ml. d) stock ammonium persulfate; 1.0 gram of ammonium persulfate was dissolved in 10 ml of ddH2O. e) TEMED stock; TEMED stock solution was supplied by the manufacturer (Biorad). Separation Minigels were prepared from solutions with a final volume of 30.12 ml that consisted of varying amounts of stock acrylamide solution and ddH2O depending on the final % desired, and unvarying amounts of stock
separating gel buffer (7.5ml), stock ammonium persulfate solution (90ul), and TEMED (30ul). Stacking portions of the minigels were prepared from solutions with a final volume of 6.71ml that consisted of 1.0ml of stock acrylamide solution, 4.0ml of ddH2O, 1.67ml of stock stacking gel buffer, 30ul of stock ammonium persulfate solution, and 10ul of TEMED stock solution. Protein samples were prepared appropriately, depending on their source (see induction below; Calzone et al., 1988) and run on the gel in 1X SDS sample buffer (Walker, J.M., 1984). Gels were ran using a chamber apparatus supplied by the manufacturer in a buffer that consisted of 0.192M glycine, 0.025m Tris pH 8.3 and 0.1% w/v SDS with a run time range between 40 and 60 minutes at 150 Volts (constant V). Proteins run on the gels were visualized by staining the gels with a solution consisting of 0.25g Coomassie Brilliant Blue R250, 125ml Methanol, 25ml Glacial acetic acid, and 100ml of ddH2O for 10 minutes and then destaining overnight in a solution consisting of 50ml Methanol, 50ml Glacial acetic acid, and 400ml of ddH2O.

**Homeobox fusion protein expression construct**

**Plasmid construct**

Plasmid vector pET3c was used to express a partial TgHBox4 genomic restriction fragment (Dolecki et al., 1988), fused to a 13 amino acid leader peptide contributed by the vector (Studier, 1986). To obtain compatible restriction enzyme sites for pET3c the TgHBox4 restriction fragment was first
put into the vector BluescriptII SK+ (Stratagene). 4ug of the TgHBox4 containing plasmid and 1ug of the Bluescript vector was digested with 5 units of Cla1 and 5 units of Pst1 in 50mM NaCl, 50mM Tris-HCl pH 8.0, 10mM MgCl2, 100ug/ml bovine serum albumin (BSA), and 5mM spermidine in a volume of 10ul by incubating at 37 degrees Celsius (C) for 1.5 hours. DNA was checked for digestion by running a TAE mini gel containing 200ng of digested DNA. The appropriate digest products were then used for a ligation reaction. The ligation mix was used to transform *E. coli* DH5 cells. Well isolated colonies were selected using LB plates with an ampicillin concentration of 100ug/ul. Colonies were screened by preparing 2ml cultures of mini-plasmid DNA, and restriction enzyme digestion with Cla1 and Pst1 was used to excise the 410 base pair insert. The pET3c vector and the Bluescript vector containing the TgHBox4 insert were prepared by digesting 5ug of each plasmid first with Cla1 in the above restriction enzyme buffer for 1.5 hours. The enzyme was then heat inactivated by incubating the digests for 10 minutes at 65 degrees C. Beta-mercaptoethanol (Sigma) was added to a final volume of 10mM along with 10 units of BamH1 and the digest was allowed to proceed at 37 degrees C for 1.5 hours. The DNA was checked for digestion by running a TAE mini gel with 200ng of digested DNA. The appropriate digestion products were then used for a ligation reaction. The ligation mix was used to transform *E. coli* DH5 cells. Well isolated colonies were selected on LB plates with an ampicillin concentration of 100ug/ul. Colonies were screened by preparing 2ml cultures
of mini-plasmid DNA, and restriction enzyme analysis with ClaI and BamHI to
cutse the 421 base pair insert. Clones that contained inserts were used to
transform BL21 Lys(S) cells (Studier 1986).

Expression of construct

Well isolated colonies were selected using M9 plates containing 100ug/ul
ampicillin and 30ug/ul chloramphenicol and used to inoculate 5ml of M9 broth
containing 100ug/ul ampicillin and grown overnight at 37 degrees Celsius. 1ml
of overnight culture was used to inoculate 100ml of M9 broth containing
100ug/ul ampicillin and 30ug/ul chloramphenicol and grown at 37 degrees
Celsius for 2.5 hours. At this point 2ml of culture was removed for use as a
control sample, and IPTG was added to the remaining culture to a
concentration of 0.4mM. After 30 minutes rifampicin (Sigma) was added to the
culture at a concentration of 200ug/ml of culture. The culture was allowed to
grow for an additional 1.5 hours at 37 degrees Celsius. Aliquots of culture
representing 8.0*10^8 cells, of the noninduced culture (no IPTG), and the
induced culture, were placed in 1.9ml centrifuge tubes. The cells were pelleted
with a 2 minute spin in a microfuge, the supernatant removed, and the pellets
rinsed with 500ul of 10mM TRIS-HCl pH 7.5. The rinsed cells were then spun
down again, the supernatant removed, and the cells resuspended in 100ul of
10mM tris-HCl pH 7.5. To this 100ul of 2X SDS sample buffer (Walker, 1984)
was added, the cells vortexed, heated to 90 degrees Celsius for 3 minutes, and
then quick chilled on ice. 5ul of each sample was electrophoresed on a 12%
Mini-PROTEAN 11 Dual Slab Gel at 150 Volts for 50 minutes in a reservoir buffer that was 0.192M glycine, 0.025 Tris pH 8.3, and 0.1%w/v SDS. After electrophoresis, the gel was stained with Coomassie brilliant blue R-250 (Walker, J.M. 1984).

Protein purification

SDS-PAGE gels

Fusion proteins were purified as follows: Whole cell lysates in 1X SDS sample buffer were electrophoresed on 4 16cm X 13cm X .3cm 10% or 12% separating SDS-PAGE gels, with 1.3cm stacking gels. They were run for one hour at 100 Volts and then for 3 to 5 hours at 150 Volts in reservoir buffer that was 0.192M glycine, 0.025M Tris pH 8.3, and 0.1%w/v SDS (1984 Walker). Volumes of sample loaded depended on the quantity of fusion protein in each induction lysate, but the range was between .750ml to 2.5ml. After electrophoresis, the gels were stained for 5 minutes in a 0.6% serva blue R solution that was prepared as follows: 3 grams of serva blue R was dissolved in 100ml of methanol and then added to 400ml of 0.1M Tris:HCl pH 7.6. The gels were then destained overnight in double distilled water, with 2 changes of the water.

Electroelutions and precipitations

The fusion protein band of each gel was excised out with a clean single edge razor. These slices were cut to fit electroelution sample cups immersed in
one buffer chamber, and the protein was electroeluted at 150 Volts for 4 hours out of the gel slices into collection cups immersed in a second buffer chamber. The electroelution buffer was 10mM Tris-Acetate pH 8.6, 0.1% SDS. The buffer in each chamber was mixed together and put back in the chambers every 30 minutes to compensate for the pH changes observed in the chambers during electroelution. Samples were removed from the collection cups every hour with a 1000ul Gilson pipette man. After electroelution, 4 sample volumes of ice cold acetone was added to the pooled collected samples and placed at -20 degrees Celsius overnight. The precipitate was pelleted by centrifugation at 9K for 30 minutes in a HB-4 rotor at 4 degrees Celsius. The pellet was rinsed with 80% acetone and then spun down again at 9K for 30 minutes. The tubes with the pellets were then put in a vacuum sealed desiccator for at least one hour and then the pellets were dissolved in 10mM Tris-HCl, pH 7.5 or PBS, pH 7.4, in a volume ranging from 500 to 1000ul. Protein concentrations of the samples were determined using the Bradford method (Bradford, M.M., 1976) and the identity of the isolated proteins were verified by electrophoresing 1ug samples on 10% or 12% acrylamide Biorad 2 mini-gels.

Production of antibodies

New Zealand rabbits were used for the production of polyclonal antibodies using the following immunization schedule: 100ug of purified protein was injected once and then 100ug was injected every 2 weeks until a total of 5
injections had been carried out. Final bleeds were tested for specific antibody content by reacting the sera to westerns (described later) containing the fusion protein that was the antigen.

**Affinity Purification of Antibodies**

**Preparation of** E. coli **lysate column**

First a one liter culture of BL21 LYS(S) E. coli cells was grown to the stationary phase. These were then recovered by pouring the culture into 4250ml Nalgene bottles, and centrifuging them for 20 minutes at 4 degrees Celsius, using a Sorvall GSA rotor, in a Sorvall RC2-B centrifuge. The broth was poured off and the bottles inverted on paper towels to remove any remaining broth. The 4 E. coli pellets were resuspended in 25ml of 0.1M sodium borate (pH 8.0), 1M NaCl and combined. 200mg of lysozyme was added to the solution and incubated for 20 minutes at room temperature. 200ul of Triton X-100 was then added and the suspension was sonicated until the viscosity decreased. The bacterial lysate was then centrifuged at 8000g's for 20 minutes. The supernatant was poured into a 50ml sterile conical tube and the pH adjusted to 9.0 with 1N NaOH. The protein concentration of the solution was determined using the Bradford method (Bradford, M.M., 1976). The proteins contained within this solution were bound to CNBr-activated Sepharose 4B (Sigma) in the following way: First 0.6 grams of CNBr-activated Sepharose 4B was washed and swelled to 2ml over a sintered glass filter with 200ml of
1mM HCl. This was mixed with 20mg of **E. coli** lysate protein at a gel:buffer ratio of 1:2. The solution was mixed on a rotary mixer overnight at 4 degrees Celsius. The next day the gel solution was transferred to a 15ml disposable centrifuge tube (Fisherbrand). The tube was centrifuged up to 500 R.P.M.'s in a refrigerated Sorvall RT6000B centrifuge using the Sorvall 1000B rotor. The supernatant was removed and replaced with 5 volumes of 1M ethanolamine or 0.2M glycine, pH 8.0 to block any remaining active groups on the gel. The protein concentration of the supernatant was determined using the Bradford method (Bradford, 1976) to approximate % of protein bound to the gel. The gel solution was mixed overnight at room temperature on a rotary shaker. After this, excess adsorbed proteins were removed by alternate rinses of the gel with 0.1M sodium borate (pH 8.0), 1M NaCl and 0.1M acetate buffer (pH 4) with .5M NaCl. This was facilitated using the centrifuge as described above. The gel was then stored at 4 degrees Celsius in 0.1M sodium borate (pH 8.0), 1M NaCl until it was needed.

**Use of the E. coli lysate column**

The gel material was spun briefly as described above and the supernatant replaced with Tris-buffered saline(TBS), (25mM Tris:HCl, 137mM NaCl, 2.7mM KCl, pH 7.4), with 0.02% sodium azide. This was repeated until the gel was equilibrated in the TBS. The antibody that was to be purified was diluted in 4ml of TBS to a theoretical concentration of 0.5mg/ml of IgG and added to the 2ml slurry of gel in a plastic scintillation vial. This was mixed overnight at room
temperature with a rotary mixer. The next day the mixture was loaded into a quik-snap column with a bottom disc (Isolab-Inc.). The solution was released from the column, used to rinse the scintillation vial, and reintroduced onto the column material. The solution was then collected in a 17*100mm culture tube (Fisherbrand). The column material was then rinsed with 5 1ml aliquots of TBS which were collected in 1.9ml microfuge tubes. The concentration and location of the antibodies in all the aliquots was determined by measuring the optical density of a sample of each aliquot at 280nm. The antibody containing aliquots were stored at 4 degrees Celsius until they were needed.

Preparation of affinity purification strips

First 2 SDS-PAGE gels were run as described in the protein purification section. The sample loaded onto the gels were E. coli cell lysate containing the antigen that the antibody to be purified was raised against. After electrophoresis, each unstained gel was placed in 0.5M KCl for approximately 15 seconds which precipitates SDS in regions where protein bands are not present. The "clear" band that represents the protein of interest was excised from the gel and placed in 50mM NaPO4, pH 6.5, 0.1% SDS. The gel slices were rinsed three times in this solution at room temperature for 30 minutes each rinse. The protein containing gel strips were then ready for transfer to aminophenylthioether (APT) cellulose paper (Schleicher and Schuell) that has been activated by diazotization. Activation of the paper was carried out as follows: First strips of APT paper were placed in cold 1.2N HCl to which had
been added 3ml of a freshly made 10mg/ml solution of sodium nitrite. This was incubated in a cold room (4 degrees Celsius) for 15-30 minutes with occasional shaking. At this point the paper had turned canary yellow. After this the strips were rinsed with 2 rinses of ice cold ddH2O and 2 rinses with 50mM NaPO4 pH 6.5. Each rinse took no longer than 5 minutes each. The protein in the gel strips was then transferred to the activated paper (diazophenylthioether-DPT) by western transfer as described below. The buffer used for the transfer was 50mM NaPO4 pH 6.5, and the transfer was carried out for 4 hours at 0.6 amps in the cold room (4 degrees Celsius). The paper was orange after transfer and the area where the protein was transferred was lighter than the surrounding area. The DPT paper was then blocked overnight using a 100mM Tris-HCl pH 9.0 solution containing 0.25% gelatin and 10% ethanolamine. After blocking, the strips were rinsed with three 30 minutes washes on a rotary mixer using 50mM Tris-HCl, pH 7.5 with 5mM EDTA and 150mM NaCl (Buffer I). The paper was then stored in Buffer I at 4 degrees Celsius until it was needed for use.

Test of affinity purification strips

To check on the efficiency of transfer, ends of the affinity strips were excised and reacted with the antibodies that were raised against the bound antigen. The appropriate sera was diluted 1/500 with Buffer I, which included 1% Bovine Serum Albumin (BSA), and reacted to one of the ends of the strips overnight. The other end was reacted similarly, but with sera that had not been exposed to antigen (pre-immune sera). The next day the ends were rinsed 3 times for
10 minutes with Buffer I and then reacted to peroxidase conjugated anti-rabbit (secondary) antibody for 2 hours. The ends were then rinsed 3 times as above, and then reacted to 4-chloro-1-naphthol and \( \text{H}_2\text{O}_2 \) at a concentration of 0.05% and 0.01% respectively, in PBS at room temperature for 35 minutes.

Use of Affinity Purification Strips

The affinity strips were cut into 8mm*8mm squares and placed in 10ml syringes with large gauge needles. The antibodies retrieved from the e. coli lysate column were mixed with Buffer I with 1% BSA to a final volume of 5ml. The effective dilution of antibody was 1/250. The syringes were immobilized on a rotary mixer and incubated overnight at room temperature. The next day the solutions were squirted out and each set of affinity squares were rinsed three times for ten minutes with 10ml of Buffer I. The antibodies were eluted with NaI and concentrated with Centricon filters (Amicon). After the last rinse, 4ml of 5M NaI made up fresh in ice water was draw up into the syringe and the solution mixed on a rotary mixer at room temperature for 8 minutes. During the elution, the upper reservoirs of two Centricon filters with a 30,000 M.W. cutoff were loaded with 300ul of Buffer I with 1% BSA. After the eight minute elution, 2ml of the elution solution was mixed with the solution in the upper reservoir of each Centricon filter. These were then spun at 5500 R.P.M.’s for thirty minutes in a Sorvall RC-5B refrigerated centrifuge at 4 degrees Celsius using an SS-34 rotor. After the spin, the liquid in the bottom reservoirs was discarded and the liquid remaining in the upper reservoirs brought to a volume of 2ml with Buffer I. The
filters were spun for 30 minutes as before. The liquid in the bottom reservoirs was again discarded and 250ul of Buffer I was added to each upper reservoir. The upper reservoirs were then capped with storage reservoirs, inverted and spun at 1000*g for 3 minutes. The storage reservoirs were then retrieved, the volumes of each solution recorded, and sodium azide (Sigma) added to each reservoir to a final concentration of 0.02%. The antibody solutions were stored at 4 degrees Celsius until further use. The affinity purification paper was kept in the syringes in Buffer 1 and stored at 4 degrees Celsius.

Pre clearing of Affinity Purified Antibodies

Preparation of sea urchin egg powder

A sea urchin culture of approximately 15-20ml of embryos was grown (described later) until the 4 cell stage. At this point the embryos were harvested and homogenized in a hand held homogenizer in 10 volumes of cold acetone (-80 degrees Celsius). The homogenate was allowed to settle and the acetone was removed and replaced to repeat the extraction process. This was repeated 2 more times and then the homogenate was transferred to a sterile 50ml conical tube. The tube was left open in a vacuum desiccator to allow the urchin powder to dry. This was stored indefinitely at -20 degrees Celsius until needed.
Primary Antibody preclearing

50mg of dry embryo powder was prewetted with PBS in 1.9ml microfuge tubes. The powder was then pulse centrifuged and the excess PBS removed and replaced with a 500ul PBS solution containing 10% NGS, .3% Triton X-100 and affinity purified antibody at the dilution appropriate for the staining of embryos (1/250 to 1/1000). This was placed at 37 degrees Celsius for 1 hour. After one hour, the mixture was centrifuged and the antibody solution removed and saved for an antibody embryo reaction.

Secondary Antibody preclearing

100mg of dry embryo powder was prewetted with PBS in a 1.9ml microfuge tube. The powder was then pulse centrifuged, the excess PBS removed and replaced with 1ml of 12mg/ml solution of Rhodamine conjugated affinity purified goat anti-rabbit antibodies (Cappel). This was put at 37 degrees Celsius for 1 hour. After one hour the mixture was spun and the antibody solution was removed and added to another 100mg of prewetted sea urchin powder. This was repeated one more time, and then the preabsorbed antibodies were stored at 4 degrees Celsius until they were needed.

Western Blots

Preparation

Mini-PROTEAN gels were run with protein samples that contained the antigens of interest for each specific western reaction. After electrophoresis,
gels were equilibrated for 5 minutes in western transfer buffer that consisted of 50mM Tris, 384mM glycine, 0.1% w/w SDS, and 20% v/w methanol. 9cm*10cm fiber pads (Biorad) and 9cm*10cm 3mm Chr Whatman paper were wetted with transfer buffer. 7.5cm*10cm .45um pore size PVDF membranes (Immobilon) were prewet with methanol for 1 minute, rinsed with ddH2O for 1 minute, and then finally wetted in transfer buffer. Western transfers were then set up on mini Trans-Blot cells (Biorad). On the negative plate electrode of each gel holder cassette the materials were layered in the following order: 1 fiber pad, 2 pieces of the Whatman filter paper, a gel for transfer, 1 piece of immobilon, 2 pieces of Whatman filter paper, and 1 fiber pad. The resulting stacks were cleared of air bubbles and then held together by securing the positive plate electrodes of the gel holder cassettes with the sliding bar and placing them in the modular electrode assembly. The modular electrode assembly was then placed in a Trans-Blot cell buffer tank that was filled with transfer buffer and contained a magnetic stir bar. The western electrophoretic transfers were then carried out at 4 degrees Celsius at a constant voltage of 100V for 40 minutes using a 200/2.0 power supply (Biorad), with constant transfer buffer mixing by utilizing a magnetic stir bar and a Magne-matic stirrer (Thomas). After transfers the Immobilon blots were removed, allowed to dry and stored at 4 degrees.
Staining

Before staining, each blot was stained with Ponceau red (Sigma), to determine efficiency of transfer and orientation of the protein samples, and to stain the molecular weight standards. Each blot was wetted with methanol, rinsed with ddH2O, and then immersed in diluted stain for 1 minute, followed by a 30 second rinse in ddH2O. The blots were dried and the molecular weight standard lanes were removed with a single edge razor blade and saved for antigen molecular weight determination, after the remaining portion of the blots were stained with antibodies. To prepare for antibody staining each blot was pre-wetted with methanol, rinsed with ddH2O, and placed in Tris buffered saline with .05% Tween-20 (TBST) for 20 minutes to rinse away the ponceau red stain. The blot would then be blocked with 1% albumin or 1% ovalbumin in TBST for 30 minutes. Finally the blot would be transferred to a solution of TBST with 1% albumin or 1% ovalbumin and the primary antibody of choice at a dilution of 1/500 to 1/5000. The primary reaction was carried out for one hour unless stated otherwise. The blot was then rinsed one time with TBST, and then 3 times for 10 minutes with the same buffer. The blot was then transferred to a solution of TBST containing 1% ovalbumin or 1% albumin and an anti-rabbit or anti-mouse IgG (whole molecule) alkaline phosphatase conjugate (Sigma) at a dilution of 1/7500. The blot was allowed to react for one hour and then was rinsed as it was after the primary antibody reaction. The blot was then developed for 5 to 15 minutes in subdued light in alkaline phosphatase buffer.
that consisted of 10mM Tris pH 8.0, 150mM NaCl, and 0.05% Tween-20 (Sigma) and contained 5mg of nitro blue tetrazolium (Grade 3) (Sigma) and 5mg of 5-bromo-4-chloro-3-indolyl phosphate (p-toluidine salt) (Sigma). The reaction was halted with a gel destaining solution containing 10% acetic acid and 10% methanol diluted in ddH2O. The blot was then rinsed with ddH2O and allowed to dry.

Rearing of Embryos

Eggs from the sea urchin *Tripnuestis gratilla* were collected by cutting out the Aristotle’s lantern of one urchin, removing the coelomic fluid and replacing it with approximately 5 ml 0.5M KCl. The urchin was then inverted over a beaker of sea water. Sperm was collected from one member of the same species by injecting approximately 2 ml of 0.5M KCl into the underside of the urchin, with the sperm being removed with a pasteur pipette and placed in a microfuge tube on ice. Eggs were filtered through a 102 micron mesh Nitex cloth and washed by settling at 1*g through two changes of MFSW (0.45micron Millipore Filtered Sea Water). After settling, the supernate was aspirated off and the eggs were suspended to approximately 1% (v/v), and fertilized with 10ml of semen (1:40) in MFSW. After fertilization, the eggs were washed as above to remove the sperm and then diluted to a final volume of .06% to .1% in TSW 8.5 [MFSW, 10mMTris-HCI pH8.5, 100micrograms Sulfadiazine (Sigma)/mL] and paddle stirred at approximately 40 RPM at 24 degrees Celsius with aeration.
The sea water was aspirated off every 24 hours through a 28micron mesh Nitex filter, and replaced with fresh TSW. Embryos were harvested at 18hrs., 24hrs., 36hrs., and 61hrs. post fertilization, corresponding to mesenchyme blastula, gastrula, prism, and pluteus stages respectively.

Harvest of Embryos

Embryos from the various stages were harvested as follows: the embryos were centrifuged in 250ml Nalgene bottles at 2000RPM's for 5 minutes in a Sorvall GSA rotor. Embryos were washed by resuspension and centrifugation through cold Acid Sea Water (ASW:0.02M Acetic acid in MFSW, pH adjusted to 4.5 with NaOH), followed by two washes with MFSW. 2mL aliquots of embryos were then transferred into 50mL conical tubes and stored at 4 degrees Celsius for no longer than 5 minutes before use.

Fixation, Blocking and Immunofluorescent Staining of Embryos

All excess MFSW was aspirated off 2mL aliquots of embryos which were subsequently fixed in 10mM Tris:HCl pH 7.6, 3.7% formaldehyde MFSW for 3 hours at room temperature. The fixer was then aspirated off and the embryos were permeabilized with cold acetone (-20 degrees Celsius) for 10 minutes. The acetone was removed and the embryos were rinsed three times with 1X phosphate buffered saline (PBS), pH 7.4. The embryos were then blocked with PBS containing 10% normal goat serum (NGS) and 0.3% Triton X-100 overnight.
at 4 degrees Celsius. The embryos were then allowed to settle and the blocking solution was removed. 100ul aliquots of these embryos were added to 500ul of PBS containing 10% NGS, 0.3% Triton X-100 and precleared affinity purified primary antibodies diluted to the appropriate concentration. These were reacted overnight at 4 degrees Celsius with constant rotation using a Labquake rotator (Labquake). Afterwords, these embryos were rinsed 3 times with PBS containing 10% NGS and 0.3% Triton-X 100. The embryos were then reacted to precleared affinity purified rhodamine conjugated goat anti-rabbit secondary antibodies that were diluted 1/300 in the same buffer for 1 hour at room temperature. The embryos were then washed 3 times in PBS containing 10% NGS and 0.3% Triton-X. 5ul of embryos were then suspended in a solution of 9:1 glycerol:PBS, mounted on microscope slides and covered with #1 coverslips elevated with vaseline "feet". A Zeiss Axio phot microscope with the appropriate filters and a 40X oil immersion lens was used for observation and photography of the embryos.
CHAPTER III

RESULTS

Expression construct production and induction

The strategy for making the TgHBox4 homeobox bacterial expression construct is shown in Figure 2. The 412 base pair homeobox containing Pst1-Cla1 fragment from a previously isolated genomic clone (Dolecki et al., 1988), was inserted into the pET3c expression vector (Rosenberg et al., 1987). First the fragment was shuttled through the pBluescript (Stratagene) vector to introduce a BamH1 site into the 5' end of the insert. The BamH1-Cla1 fragment was then cloned into the pET3c vector. The completed construct was used to transform competent DH-5 E. coli cells. A positive colony was picked from ampicillin plates, grown up, and the vector DNA isolated by plasmid miniprep. This DNA was then used to transform BL21 E. coli cells that include plysS, a T7 lysozyme producing vector that also confers chloramphenicol resistance (Rosenberg et al., 1987). The lysozyme producing vector suppresses any transcription of the expression vector prior to induction. A positive colony obtained from ampicillin-chloramphenicol plates was grown up, and the resultant culture induced with IPTG to produce the homeodomain containing
Figure 2. TgHBox4 expression construct.

Strategy for the formation of the TgHBox4 expression construct. The pGEMHBox4 and pBluescript II SK +/- vectors were digested with Pst1 and Cla1. The TgHBox4 Pst1-Cla1 fragment was isolated and ligated into the digested pBluescript II SK +/- vector to give the clone pBluescript II Hbox4.

The pET-3C expression vector and the pBluescript II Hbox4 vector were digested with BamH1 and Cla1. The TgHBox4 fragment was isolated and ligated into the pET-3C vector to yield the expression construct pET-TgHBox4. The BamH1 site obtained from the pBluescript vector allowed the cloning of the TgHBox4 fragment into the pET-3C expression vector in the correct reading frame.
protein. Success of induction was determined by running out aliquots of bacterial lysate on 12% SDS-PAGE mini-gels such as the one displayed in Figure 3a. Lane 1 contains bacterial lysate prepared from cells before the addition of IPTG. Lane 2 and 4 is bacterial lysate prepared from cells two hours after induction by the addition of IPTG. What can be seen is the appearance of a band in the 14kd range. This was higher than than ~12kd predicted from the open reading frame that included a short sequence from the pET vector itself.

To verify the protein being expressed was encoded by the TgHBox4 fragment, an antibody raised against a synthetic peptide representing part of the protein sequence encoded by the TgHBox4 nucleotide sequence was reacted to a western blot that included the expressed protein. The result is shown in Figure 3b + 3c. They are two identical western blots. The first one (c) was reacted to pre-immune rabbit sera and the second one (d) was reacted to the anti-HB4 peptide sera. Lane 1 contains bacterial lysate prepared from cells prior to the addition of IPTG. Lane 2 and 4 contain bacterial lysate prepared from cells two hours after induction with IPTG. The synthetic peptide antisera reacted strongly to the IPTG induced band of bacterially expressed protein, indicating that it represents a protein encoded by the TgHBox4 sequence.
Figure 3. Characterization of the bacterial expression protein.

Figure 3A is a 12% acrylamide gel stained with coomassie brilliant blue. Lane 1 contains bacterial lysate from TgHBox4 construct containing *E. coli* cells prior to the addition of IPTG for the induction of expression. Lanes 2 and 4 contain bacterial lysate from TgHBox4 construct containing *E. coli* cells 2 hours after the addition of IPTG for the induction of expression. Lane three is a sample of TgHBox4 bacterially expressed protein that was gel purified. Figure 3B and 3C include western transfers of gels that were replicates of the gel represented in Figure 3A. The blot shown in Figure 3B was reacted to pre-immune rabbit sera. The blot shown in Figure 3C was reacted to sera from a rabbit immunized with a TgHBox4 encoded synthetic peptide. Figure 3C displays strong antibody antigen reactions where the TgHBox4 bacterial protein is present on the western blot.
Protein purification and antibody production and purification

Once the 14kd protein was identified as a product of the TgHBox4 expression vector, it was purified and used to immunize rabbits to produce polyclonal antibodies. Aliquots of bacterial lysates were run out on large SDS-PAGE gels, the TgHBoxHB4 protein bands excised, and the protein electroeluted out of the gel fragments. The protein was then precipitated in acetone, resuspended in 10mM tris buffer, pH 7.5 and an aliquot ran out on an SDS-PAGE minigel to determine it’s concentration and purity. Lane 3 of Figure 3a displays the result of a purification. Lane 3 of Figure 3c shows the reaction of anti-HB4 peptide sera to the purified protein. The purified protein was used to immunize a rabbit to produce antibodies against the expressed protein (anti-TgHBox4). The success of antibody production was determined by reacting the sera from the injected rabbit to a western blot that included bacterial lysate prepared from uninduced and induced cells (see Figure 4A). Lanes 1 through 4 contain bacterial lysates prepared from uninduced and induced cells that contain expression vectors for another homeotic protein, TgHBox1. Lanes 5 and 6 contain bacterial lysates that were prepared from uninduced and induced cells containing the TgHBox4 expression vector, respectively. Lane 6 shows a strong antibody reaction to the TgHBox4 expressed protein.

The specificity of the anti-TgHBox4 sera was tested by reacting it to a western strip that an ovalbumin-HB4 peptide had been transferred to (Figure 5). The first strip in the figure was reacted to pre-immune sera (18-0). The strip to
Figure 4. Characterization of the anti-TgHBox4 expressed protein antibodies.

Reaction to *E. coli* lysate.

Figure 4 is a western transfer with various expression lysates bound to the nitrocellulose support. The blot has been reacted to anti-TgHBox4 sera. Lane 1 and lane 2 contain bacterial lysates from *E. coli* cells that contain the expression vector TgHBox1cDNA. Lane three and lane four contain bacterial lysates from *E. coli* cells that contain the expression vector TgHBox1genomic. Lane 5 and 6 contain bacterial lysates from *E. coli* cells that contain the expression vector pETHBox4. The lysates in lanes 1, 3 and 5 were prepared prior to induction with IPTG, whereas the lysates in lanes 2, 4, and 6 were prepared 2 hours after induction with IPTG. Lanes 2 and 4 show the antibodies displaying a little cross reactivity to the homeobox containing expressed proteins. Lane 6 shows the antisera reacting strongly to the bacterially expressed TgHBox4 protein used to produce it.
Figure 5. Characterization of the anti-TgHBox4 expressed protein antibodies.

Reaction to a TgHBox4 synthetic peptide.

Figure 5 includes two western transfer strips. They both have a synthetic TgHBox4 peptide albumin conjugate bound to them. The strip on the left was reacted to preimmune rabbit sera. The strip on the right was reacted to anti-TgHBox4 sera. The sera shows a strong reaction to the HB4 peptide, while no reaction is seen with the pre-immune sera from the same rabbit.
the right was reacted to the anti-TgHBox4 sera and a strong reaction is apparent.

The anti-TgHBox4 antibodies were affinity purified with TgHBox expressed protein immobilized on APT strips. The success of the affinity purifications were tested by reacting the affinity purified antibodies to western blots containing uninduced and induced E. coli TgHBox4 expression lysates (Figure 6). Figure 6A is a western blot reacted to unpurified anti-TgHBox4 sera and and Figure 6B is a western blot reacted to affinity purified anti-TgHBox4 antibodies. Each lane 1 contains uninduced lysate and each lane 2 contains induced lysate. The arrows point to the reaction of anti-TgHBox4 antibodies to the protein that they were raised against. A significant drop in non specific backround staining is apparent when comparing Figure 6B to Figure 6A.

Immunofluorescent localization of TgHBox4 expression

Localization determination of TgHBox4 proteins in developing embryos was carried out using anti-TgHBox4 affinity purified antibodies. First embryos were harvested at the three developmentally important stages of blastula, gastrula, and pluteus. The embryos were washed in cold acid sea water, rinsed in MFSW, fixed with formaldehyde, blocked, and rinsed thoroughly. The embryos were incubated with anti-TgHBox4 antibodies that were diluted and preabsorbed with 4 cell stage sea urchin egg powder to prevent nonspecific antibody binding. The localization of the primary antibodies was detected with
Figure 6 Affinity purification of the anti-TgHBox4 expressed protein antibody.

Figure A and B are two western blots. Both have the same proteins of lysates from *E. coli* containing the expression vector TgHBox4. Lane one contains lysate from cells before exposure to IPTG. Lane 2 contains lysate from cells 2 hours after the induction of the TgHBox4 expression construct with IPTG. Blot A has been reacted to whole TgHBox4 anti-sera. Blot B has been reacted to anti-TgHBox4 antibodies that have been affinity purified against purified TgHBox4 protein bound to APT strips. The arrows in lanes A2 and B2 indicate the protein band representing the antigen used to immunize the rabbits. Figure 6B shows a significant reduction of non-specific background antibody reactions while maintaining the strong specific antibody-antigen reaction.
rhodamine conjugated goat anti-rabbit antibodies. The stained embryos were viewed at 40X with an axiophot microscope and photographed. Figure 7A is a photograph of the immunolocalization of TgHBox4 protein in the blastula stage embryo. The staining appears generalized with some localization in the two corners of the ventral side where the triiradiate spicules will develop from the primary mesenchyme cells. Figure 7B is a photograph of the immunolocalization of TgHBox4 protein in a gastrula stage embryos. The embryo is shown with the ventral side facing the viewer. The protein expression is still generalized with some localization defined by the arrangement of the primary mesenchyme cells and in the gut region. The Figure 7C and 7D are photographs of the immunolocalization of TgHBox4 protein in pluteus stage embryos. The embryos in Figures 7C and 7D are at early and late pluteus stage, respectively. The staining pattern is extremely localized at this stage to the posterior region of the large intestine.

Isolation and characterization of intestinal cDNA clone

A Pst1-Cla1 412 base pair fragment from a previously isolated TgHBox4 genomic clone (Dolecki et al., 1988), was randomly labeled, and used to screen a Tripneustes gratilla intestinal cDNA library made in lambda zap (Stratagene) by Marcus Noll. After screening 200,000 recombinants, one positive clone was detected. A positive plaque of this clone was isolated, and it's insert isolated in a plasmid using the in vivo excision method (Stratagene).
Figure 7. Spatial localization of TgHBox4 proteins during embryogenesis.

The embryos of each stage were fixed with formaldehyde, and reacted to sequentially affinity purified polyclonal anti-TgHBox4 antibodies and rhodamine labeled goat anti-rabbit secondary antibodies. A is a blastula stage embryo. There is generalized antibody staining, with some localization on the ventral side where the triiradiate spicules will form out of two groups of primary mesenchyme cells. B is a gastrula stage embryo with the ventral side toward the viewer. Staining is generalized with some localization in the regions where the primary mesenchyme cells are located and in the gut region. C and D are posterior views of embryos at earlier and later pluteus stage, respectively. Antibody staining is now localized to the large intestine.
The "phagemid" DNA was isolated and amplified by transforming XL-1 blue host cells, picking a colony from ampicillin plates, growing the bacteria in LB broth and performing a plasmid mini-prep on the overnight culture. The plasmid DNA was initially characterized using restriction digests and agarose gels. Next the clone was sequenced. Both strands representing the region 5' prime of the initiation codon and the open reading frame were sequenced. One strand representing the remainder of the clone, including the poly-A signal and site were sequenced. The length of the clone was found to be 3634 base pairs (see Appendix C), in close agreement with the size of the TgHBox4 mRNA found in the intestine, as determined by Northern blot analysis (Dolecki et al., 1988).

Sequence comparison of the TgHBox4 clone confirmed the location of a putative splice site identified in the previously isolated genomic clone gTgHBox4-1 (Dolecki et al., 1988). Figure 8 displays the alignment of the Aat11-EcoR1 fragment of the cDNA clone with the homeobox containing Hind11-EcoR1 fragment of the genomic clone. The two sequences are identical from the EcoR1 sites to just 5' of the Mbo11 sites. This area in the genomic clone is characteristic of an intron exon splice site (Shapiro and Senapathy, 1987.) and the divergence of the two sequences at this point confirms a splice junction.

The sequence from the first initiation codon to the first stop codon was translated using PCGENE. This analysis yields a putative protein with a size of 32,321 kilodaltons (kd) (see Figure 9). This sequence includes two areas that
Figure 8. Alignment of homeobox containing restriction fragments of TgHBox4 intestinal cDNA and TgHBox4 homeobox containing genomic DNA.

The top sequence is of a Aat11-EcoR1 TgHBox4cDNA restriction fragment. The bottom sequence is of a Hind11-EcoR1 TgHBox4 genomic fragment. The alignment in the 3' to 5' direction is 100% identical upto nucleotide 827 of the cDNA and upto nucleotide 321 of the genomic DNA. This region of divergence confirms the presence of a splice site in the genomic DNA in this region.
Figure 9. Translation of open reading frame of TgHBox4 intestinal cDNA.

Translation from the first initiation codon (triangle) results in a putative protein with the molecular weight of 32,321 kilodaltons (kd). Residues noted which are found in the upstream region of Abd-B and its homolog proteins are the Aspartic acid-Phenylalanine consensus (diamonds), and the Proline-Tyrosine consensus (stars). The homeobox region of the protein is underlined.
loosely correspond to two consensus sequences found in the 5 prime region of Abd-B and its homologs. The first consensus is DXnF, where D is between 15 and 25 amino acids from the translation initiation and F is 7 amino acids or less downstream of D. The TgHBox4 sequence has a D 6 amino acids downstream of the leader methionine with F 10 amino acids from D. The second consensus sequence is P and Y found adjacent to each other between 70 and 90 amino acids from the leader methionine. The TgHBox4 sequence has a P 79 amino acids from the leader sequence with a Y 7 amino acids downstream of the P. A final area of interest in this sequence is encoded by base pairs 745 through 855. This area has a high percentage of acidic residues (30%), which is similar to the human Abd-B homolog HOX4E (Redline et al., 1992).

A Southern blot was made using the TgHBox4 homeobox containing genomic clone DNA for reaction with TgHBox4 cDNA probes. EcoR1, HindIII, and EcoR1-HindIII digests of the genomic clone were run out on an agarose gel, and transferred to nitrocellulose. The blot was separately reacted to labeled cDNA restriction fragments that represented the 3', middle, and 5' regions of the cDNA clone. The results of the reactions are displayed in Figure 10. Figure 10a is a photograph of the ethidium bromide stained DNA. Lane 1 is the 14.1 kb TgHBox4 genomic clone digested with HindIII and EcoR1. Lane 2 is the same clone digested with just HindIII. Lane 3 is the same clone digested with just EcoR1. Lane 4 is the uncut cDNA clone. Lane 5 is marker lambda DNA digested with HindIII. Figure 10b is an autoradiograph
Figure 10. Southern blots of genomic DNA clone gTgHBox4-1 reacted to cTgHBox4 intestinal probes.

A is an ethidium bromide stained gel including various digestions of the 14.1 kb lambda TgHBox4 genomic clone before transfer. Lane 1 is the 14.1 kb clone digested with EcoR1 and Hind111. Lane 2 is the same clone digested with Hind111. Lane 3 is the same clone digested with EcoR1. Lane 4 is the uncut cTgHBox4int cDNA intestinal clone. Lane 5 is lambda DNA cut with Hind111 for a size marker. B is an autoradiograph of the Southern blot reacted to an intestinal cDNA probe that includes the homeobox as well as 450 base pairs 5' of the homeobox. C is an autoradiograph of the Southern blot reacted to a intestinal cDNA probe that is an EcoR1 restriction fragment found in the middle of the intestinal clone. D is an autoradiograph of the Southern blot reacted to a 830 base pair intestinal cDNA probe that includes the 3' end of the intestinal cDNA clone.
of the Southern blot reacted to the EcoRV-EcoRV 1.1kb TgHBox4 cDNA fragment of the 5’ portion of the cDNA that includes the homeobox and ~450 base pairs upstream of the homeobox. The probe reacts to a 1.2kb Hind111-EcoR1 fragment that includes the homeobox, a large Hind111 fragment that includes one of the lambda arms, and a 1.2kb EcoR1 fragment that also includes the homeobox. Figure 10c is an autoradiograph of the Southern blot reacted to the middle EcoR1-EcoR1 0.8kb TgHBox4 cDNA fragment. The probe reacts to a 0.8kb Hind111 fragment, to the same large Hind111 fragment as the EcoRV-EcoRV probe, and a 0.8kb EcoR1-EcoR1 fragment.

Figure 10d is an autoradiograph of the Southern blot reacted to a 0.830kb Xho1-Xho1 TgHBox4 cDNA fragment that includes the most 3’ end of the cDNA clone. The probe reacts to a 1.5kb Hind111-EcoR1 fragment, to the same large Hind111 fragment as the other two probes, and a 1.4kb EcoR1-EcoR1 fragment.

The results presented above allowed the alignment of the TgHBox4 cDNA clone and the reactive fragments of the genomic clone, and the alignment is displayed in Figure 11. The genomic DNA is positioned at the top of the figure, with the cDNA below it. cDNA sequence more 5’ of the splice junction did not appear in the genomic clone which is represented by sequence divergence in Figure 8 and shown in Figure 11. Partial sequencing of the 1.4kb genomic fragment that reacted to the Xho1-Xho1 TgHBox4 probe revealed two poly-A signals.
Figure 11. Alignment of the TgHBox4 homeobox containing genomic clone with the TgHBox4 cDNA intestinal clone.

Analysis of the results displayed in Figures 7 and 8 resulted in this alignment. Partial sequence analysis of the genomic EcoR1 fragment that reacted to a 3' cDNA intestinal clone resulted in the discovery of the second poly-A signal as indicated in the Figure. The boxed region represents the location of the homeobox. Legend: HindIII; Mbol; EcoR1; splice; junction.
HB4 genomic DNA

splt. site
H III M II R I R I poly-A signals

HB4 intestinal cDNA

splt. jxn.
R I M II R I R I poly-A signal

1 kb.
One signal aligns with the one used by the intestinal mRNA. The second one is found 221 base pairs downstream from the first signal and could account for the 200 base pair difference between the mRNA found in gastrula and pluteus embryonic stages and the mRNA found in adult tissues.

**Isolation and characterization of an upstream exon containing genomic clone**

In order to characterize the TgHBox4 gene more completely, a genomic clone was sought which contained an exon representing 5' prime sequence of the intestinal TgHBox4 cDNA. An EcoRV 445 base pair fragment from the 5' prime end of the 3cTgHBox4-int intestinal clone was randomly labeled and used to screen plaques from an EcoR1 *Tripneustes gratilla* genomic library (Stratagene). A single positive clone, gTgHBox4-2 was found after 210,000 recombinant plaques were screened. The positive plaque was isolated, amplified and its DNA purified. By PCR (Figure 12) and restriction digest analysis it was determined that gTgHBox4-2 contained an exon bearing insert of 4.5kb. Sequence analysis revealed an exon that encodes all of the 5' intestinal mRNA sequence not represented by the homeobox containing exon (Figure 13). The TgHBox4 genomic clone also includes a splice site. Immediately 5' of this splice site the sequence aligns with the TgHBox4 intestinal cDNA beginning where exon sequence in gTgHBox4-1 diverged from the cDNA sequence.
Figure 12. Sizing of Upstream Exon Containing Genomic TgHBox4 Clone with PCR Analysis.

Four PCR reactions were carried out. Two reactions were carried out with the lambda primer #1218 and either a sense or antisense TgHBox4 gene specific primer. The other two reactions were carried out with the lambda #1222 primer and either a sense or antisense TgHBox4 gene specific primer. Two reactions yielded products that are shown in Figure 12 A. Lane one displays the product from the reaction with the lambda primer #1222 and the sense TgHBox4 primer #92275. Lane 2 displays the product from the reaction with the lambda primer #1218 and the anti-sense TgHBox4 primer #92117. Figure 2 B shows the position and orientation of each primer on the genomic clone (1) and the resultant products from their respective PCR reactions.
Figure 13. Alignment of 5' restriction fragment of TgHBox4 intestinal cDNA with TgHBox4 clone containing upstream genomic DNA.

The upper sequence is part of the 5' exon containing genomic TgHBox4 clone including the TATA box. The lower sequence is TgHBox4 intestinal cDNA from nucleotide number 1 to a Bgl1 site at position 860. The alignment in the 5' to 3' direction is almost 100% identical up to nucleotide 937 of the genomic sequence and nucleotide 827 of the cDNA sequence. This region of divergence confirms the presence of a splice site in the genomic DNA in this region.
The character to show that two aligned residues are identical is ‘:’

NHB4GEN       GTCTGTTGGGTTTATCACCCTCTCCACCGGACTAGCCGCTATAT -50

TATA BOX

NHB4GEN     - CGACGACCATAAAATACGAGCGAATTTAAATGACGACATCAACCTTCT -100
cDNA start site

NHB4GEN     - GTGTGATTTCTGATAAAATCGAGGATACGTGACTCCACCGTA -150
            ATTGATAAAATTCGGAGGATAACGTGACTCTCTGCTGCCTTA -200
            ::::: :::::::::::::::::::::::::::::::::::::::::::
            HB4CR1BGL1- ATTGATAAAATTCGGAGGATAACGTGACTCTCTGCTGCCTTA -40

NHB4GEN     - AAAGGCGTGTCCACCAGATTTACACTCTTCTATTCATGGATCATC -200
            ::::: :::::::::::::::::::::::::::::::::::::::::::
            HB4CR1BGL1- AAAGGCGTGTCCACCAGATTTACACTCTTCTATTCATGGATCATC -250

NHB4GEN     - GAATAACCGACGAAATTTCGAGATAGACCGGCTGGCGAGGCCTA -250
            ::::::::: ::::: :::::::::::::::::::::::::::
            HB4CR1BGL1- GAATAACCGACGAAATTTCGAGATAGACCGGCTGGCGAGGCCTA -300

NHB4GEN     - GTGTGAGAGCCGCAAATATGTATCATCTCATACAGATCAAAATACCAG -250
            ::::::::: ::::: :::::::::::::::::::::::::::
            HB4CR1BGL1- GTGTGAGAGCCGCAAATATGTATCATCTCATACAGATCAAAATACCAG -350

NHB4GEN     - CTATGTGCACTGGCCCTTGGTTTTATCCTGCCGATCACCATACCTACTCTACCA -350
            ::::::::: ::::: :::::::::::::::::::::::::::
            HB4CR1BGL1- CTATGTGCACTGGCCCTTGGTTTTATCCTGCCGATCACCATACCTACTCTACCA -400

NHB4GEN     - TGCTGATGACCGAATGGATAAGCTGGCGACGACTTCGACAACTCGGCGTA -400
            ::::::::: ::::: :::::::::::::::::::::::::::
            HB4CR1BGL1- TGCTGATGACCGAATGGATAAGCTGGCGACGACTTCGACAACTCGGCGTA -450

NHB4GEN     - TCGTACGCCATGGGAGGGCCGAAAACTTTGAATCGGGACTAATGGCGGC -450
            ::::::::: ::::: :::::::::::::::::::::::::::
            HB4CR1BGL1- TCGTACGCCATGGGAGGGCCGAAAACTTTGAATCGGGACTAATGGCGGC -500

NHB4GEN     - TTCTGAGGCTGGGATTTATCTCGATCGGTGCCAGCTATGAAACGAT -500
            ::::::::: ::::: :::::::::::::::::::::::::::
            HB4CR1BGL1- TTCTGAGGCTGGGATTTATCTCGATCGGTGCCAGCTATGAAACGAT -550

NHB4GEN     - ATCCTAATCATGCTAATCTCTATGTACCATCCTCTGTGGAGCAGGTGCTA -550
            ::::::::: ::::: :::::::::::::::::::::::::::
            HB4CR1BGL1- ATCCTAATCATGCTAATCTCTATGTACCATCCTCTGTGGAGCAGGTGCTA -600

NHB4GEN     - ATCCTAATCATGCTAATCTCTATGTACCATCCTCTGTGGAGCAGGTGCTA -600
            ::::::::: ::::: :::::::::::::::::::::::::::
            HB4CR1BGL1- ATCCTAATCATGCTAATCTCTATGTACCATCCTCTGTGGAGCAGGTGCTA -640
NHB4GEN - CACACGGGACCCATACCTGTAAGCCTAACCGCTGCCGGGACTCTGGCGAC -650
HB4CR1BGL1 - CACACGGGACCCATACCTGTAAGCCTAACCGCTGCCGGGACTCTGGCGAC -540
NHB4GEN - GAAATACATACCACTGACGGTGGTGGTGGAAACACAAGCCCGACTGGCGT -700
HB4CR1BGL1 - GAAATACATACCACTGACGGTGGTGGTGGAAACACAAGCCCGACTGGCGT -590
NHB4GEN - CAACACACACCTTGCTTGACGAAAAGCACAAATACATCGCCAGTGGCGGT -750
HB4CR1BGL1 - CAACACACACCTTGCTTGACGAAAAGCACAAATACATCGCCAGTGGCGGT -640
NHB4GEN - TATGAGGGATACAGCTGCCCGAAAGAGCCCCTAGCGACTGTCAAATCGAC -800
HB4CR1BGL1 - TATGAGGGATACAGCTGCCCGAAAGAGCCCCTAGCGACTGTCAAATCGAC -690
NHB4GEN - GGGGCAGAGAACGGTGGAGATCAGTCCTACGTACCCGTCGACGGGCTCGG -850
HB4CR1BGL1 - GGGGCAGAGAACGGTGGAGATCAGTCCTACGTACCCGTCGACGGGCTCGG -740
NHB4GEN - CCGGTGGTTCACCTATTACAGTCATCCATAATAAGGACTCGCCTGAATCA -900
HB4CR1BGL1 - CCGGTGGTTCACCTATTACAGTCATCCATAATAAGGACTCGCCTGAATCA -790
NHB4GEN - GACATGAAAGAAAATGACGTCATTGATGACTGTGATGGTAAGATGTGTTC -950
HB4CR1BGL1 - GACATGAAAGAAAATGACGTCATTGATGACTGTGATGGTAAGATGTGTTC --836
NHB4GEN - AGCAGAAGAACGGGG-ACA-CG-CCAAC------CTGGC
HB4CR1BGL1 - AGCAGAAGAACGGGG-ACA-CG-CCAAC------CTGGC -866
The open reading frame of the cDNA is continued in the gTgHBox4-2 clone 5’ of the cDNA encoding region up through another methionine codon (Figure 14). Upstream of this methionine, three stop codons are present as well as a SP-1 binding site (GC box). Comparison of the TgHBox4 intestinal cDNA sequence with the TgHBox4 genomic sequence resulted in the alignment shown in Figure 15.
Figure 14  5' region of upstream genomic clone gTgHBox4-2 including potential open reading frame.

Displayed in this figure is a translation start site that is upstream and in frame with the ORF of the TgHBox4 cDNA clone. Also included is an SP1 binding site (GC box) and various stop codons.
A 5' REGION OF UPSTREAM TgHBox4 GENOMIC CLONE

10  20  30  40  50  60
GGATCTGATTATTACCAATTGTGTGGAGGAAATAGAGGGGTAAGGCTTGAT
AspLeuIleIleThrAsnCysValPheValGluArgLysIleGluGly—GlyLeuMET

GC BOX 70  80  90  100  110  120
GAGGGCCCGTTATGGCCNAGTTGGATCACGTGACAACTTCAAGGTTCTCACTCTCATCC
ArgAlaGlyTyrTrpXAAserTrpIleThr—GlnLeuGlnGlySerHisSerHisPro

130  140  150  160  170  180
TTATGTTTTATGCCCGCAGGTCGTGGGGGTTTATCACCGTCTCCACCGACTAGCCCGTA
TyrValLeuCysProGlnValValGlyValTyrHisArgLeuHisArgThrSerPro—

190  200  210  220  230  240
ATATCGACGACATAAAAATACGAGCGATATTAAAATGACGACATCAACCTTCTGTGTGAA
TyrArgArgHisLysAsnThrSerAspIleLysMETThrThrThrPheCysValAsn

cDNA START SITE
250  260  270  280  290  300
TTCATTGATTTTGGAGGAGATAACGTGACTGCAGCTGCCCTGGCCGTGTCCCAGC
SerLeuIleAsnGlyGlyAspAsnValThrAlaAlaAlaLeuLysGlyArgValProAla

310  320  330  340  350  360
AGATTACACTCTCTATCATATGATCATGCAATTGCTCTTGACCTCCGAAATTCCG
AspLeuHisSerSerIleIleHisGlySerSerLysCysSerLeuThrProLysPheAla

370  380  390  400  410  420
CGATACGTCATCCGGCGAGGCTGCTGGTGAGGCGCGACCCACGGCAATGGACTGCG
AspThrSerIleGlnAlaThrAlaSerValValThrAlaThrGlyAsnGlyLeuAla

430  440  450  460  470  480
GACGACCCCTCAGGTGTGCGACCCGCAAAATATGTATCATCTACAAAACAGATCAAATAC
ThrThrProGlnValSerThrAlaAlaAsnMETTyrHisLeuGlnThrAspGlnAsnThr
Figure 15. Alignment of TgHBox4 exons with TgHBox4 intestinal cDNA.

Sequence comparison of TgHBox4 intestinal cDNA with the two exon containing TgHBox4 genomic clones resulted in the alignment displayed here.
TgHBox4 genomic DNA

TgHBox4 intestinal cDNA
CHAPTER 4
DISCUSSION

The homeobox containing gene, TgHBox4, found in the sea urchin species *Tripneustes gratilla*, was characterized in this study. Previous comparisons of the TgHBox4 homeodomain sequence and homeodomain sequences found in proteins in *Drosophila*, mice and frogs led to the conclusion that the TgHBox4 gene was a homolog of the *Drosophila* Abdominal-B (Abd-B) gene. The TgHBox4 homeodomain sequence was found to be 73% similar to the *Drosophila* Abd-B gene, which was the highest similarity found among all the homeobox genes of the HOM-C cluster (Dolecki et al., 1988). Based on sequence comparisons of the TgHBox4 homeodomain with homeodomain sequences among the mice and frogs TgHBox4 was aligned with genes that had been classified previously as Abd-B homologs, based on their homeodomains high sequence similarity with the *Drosophila* Abd-B homeodomain. The homeodomain of TgHBox4 was found to be 90%, 88%, and 87% similar to the homeodomains of the mice genes Hox 1.7, Hox 3.2, and the frog gene XlHbox6, respectively (Dolecki et al., 1988).

Comparisons of sequences from the homeodomain of TgHBox4 to the sequences of more recently discovered homeodomain containing proteins was carried out to corroborate the previously held hypothesis that TgHBox4 is a
Drosophila Abd-B homolog. The comparisons found the homeodomains of the genes Hox 2.5 and Hox 4.4 to be the most similar to the homeodomain of the TgHBox4 gene. Both were found to be 88% similar. Hox 2.5 and Hox 4.4 are located on different chromosomes than Hox 1.7 and Hox 3.2 and have been determined to be homologs of and align with the Drosophila Abd-B gene (Bogarad et al., 1989; Duboule and Dolle, 1989). The TgHBox4 homeodomain was also found to be similar to the Chicken Hox 4.4 homeodomain, with a similarity of 87%. This gene has also been aligned with the Drosophila Abd-B gene and its homologs. The sequence comparisons discussed support the idea that TgHBox4 is truly a Drosophila Abd-B homolog.

There are consensus sequences found on the N-terminal side of the homeodomain in the protein product of the Abd-B gene and its homologs that are unique to these proteins. Versions of these consensus sequences have been found in the protein sequence translated from the ORF of the cTgHBox4-int clone. One consensus is an aspartic acid 15 to 20 amino acids downstream of the initiation methionine followed by a phenylalanine within 7 amino acids. The second consensus is a proline 70 to 90 amino acids downstream of the initiation methionine flanked by a tyrosine (Izpisua-Belmonte et al, 1991). One area within the N-terminal region of the protein coded for by c-TgHBox4-int clone aligns loosely with the first consensus sequence (Figure 9). An aspartic acid is found 6 amino acids downstream of the first methionine with a phenylalanine found 10 amino acids downstream of the aspartic acid. The
proline tyrosine consensus is also found within the c-TgHBox4-int protein sequence (Figure 9). 78 amino acids downstream of the first methionine there is a proline with a tyrosine 7 amino acids downstream of the proline. There is also a tyrosine flanking the proline on the N-terminal side. In Abd-B and the vertebrate homologs the proline is immediately flanked downstream by the tyrosine. The function of these two amino acids are unknown however and so the orientation of the residues may not affect function. The tyrosine flanking the proline on the N-terminal side may be the tyrosine that is involved in whatever function this consensus may have. The appearance of these Abd-B consensus regions in the c-TgHBox4-int protein sequence reinforce the idea that TgHBox4 is a homolog of the Drosophila Abd-B gene.

A partial molecular characterization of the TgHBox4 gene was carried out using a previously isolated genomic clone, a newly isolated intestinal cDNA clone, and a newly isolated genomic clone.

Previous Northern blot analysis using a homeobox containing genomic fragment as a probe suggested TgHBox4 to be a temporally regulated gene, capable of producing multiple transcripts (Dolecki et al., 1988). A total of 3 different transcripts are produced in the embryo and in adult tissue. A 4.4kb transcript appears at blastula stage and persists through the gastrula stage and disappears by the pluteus stage. A 3.8kb transcript barely detectable at the blastula stage increases to a level equal to the 4.4kb message at the gastrula stage and persists through the pluteus stage. In certain adult tissues a sole
transcript of 3.6kb appears at levels at least 10 times higher than those of the messages during embryogenesis.

The high levels of message found in the adult large intestine tissue made it the prime choice as a source of mRNA. A cDNA library was made by Marcus Knoll from mRNA isolated by Gregory Dolecki from this tissue. I screened his library using the same homeobox containing fragment that was used to screen the Northern blot described above. A cDNA clone was detected, isolated and sequenced (see Appendix C), and subsequently found to be 3.6kb in size.

It was determined that this clone had a 1.1kb open reading frame that included the homeobox and encoded a 32,321kd protein (see Figure 9). This putative protein begins at the first methionine of the open reading frame. The region adjacent and including the region coding for the initiation methionine are in agreement with the Kozak consensus sequence which is found at transcription initiation sites (Kozak, M., 1991).

There is an acidic region within the sequence of the cTgHBox4-int protein which may have functional implications. 38% of the amino acids in the region from amino acid 174 to amino acid 214 within the cTgHBox4-int protein sequence are acidic. This region starts 10 amino acids upstream of the homeodomain. Protein domains that are highly acidic is a characteristic of one of the three consensus motifs of activation domains of eukaryotic transcription factors (Mitchell and Tjian, 1989). It is theorized that acidic regions like this
interact with components of the initiation complex, like for example TFIID, or RNA polymerase II.

Alignment of this new cDNA clone with gTgHBox4-1, the original genomic clone revealed that the cDNA sequence from base pair 830 (51 base pairs 5' of the homeobox) to the 3' end is encoded by one exon (see Figure 11). Additional sequence analysis of the genomic clone downstream of the cDNA coding region revealed a second poly-A signal 231 base pairs away from the poly-A signal utilized during production of the intestinal mRNA.

To further elucidate the molecular structure of TgHBox4 gene the 5' end of the intestinal cDNA clone was used to screen a genomic library. A genomic clone of approximately 4.6kb was isolated. Sequence analysis revealed that it contained an exon that included the remaining 5' coding sequence of the TgHBox4 intestinal cDNA (see Figure 13). 63 base pairs from the 3' end of the genomic clone gTgHBox4-2 the sequence begins to align with the intestinal clone cTgHBox4-int where the gTgHBox4-1 clone alignment ended. Comparison of the gTgHBox4-2 sequence where the alignment begins, with the consensus sequence for a splice site (Shapiro and Senapathy, 1987), suggests it is an exon intron junction. The alignment of the gTgHBox4-2 genomic clone with the cTgHBox4-int clone continues all the way to the 5' end of the cTgHBox4-int clone. The gTgHBox4-2 genomic clone contains a TATA box consensus (Lewin, 1990) 36 base pairs upstream of the transcription start site for the cTgHBox4-int clone. 5' of the end of the open reading frame are 3
stop codons and an SP-1 binding site. SP-1 is a transcription factor and so its presence suggests this is a promotor region.

Closer inspection of the genomic sequence reveals a possible mechanism for production of the mRNA message found in embryonic tissue that is slightly larger than the intestinal message. The region of the gTgHBox4-2 clone 5' of the intestinal cDNA encoding sequence continues to be in frame with the cDNA coding region up through another methionine initiation codon. Detailed studies of transcription initiation has led to the discovery of some genes that produce two versions of the same protein by producing two different messages (Kozak, 1991). One message includes all the sequence of a shorter transcript, plus additional sequence added to the 5' end that includes the codon for another methionine. The use of alternate promotors determines which initiation of transcription site is used and therefore which message is produced (Kozak, 1991). An example of a nuclear protein that uses this type of mechanism to produce 2 messages that code for similar but functionally distinct proteins from the same exons is the human progesterone receptor gene (Kastner et. al, 1990). It is possible that the TgHBox4 gene uses a similar mechanism to produce the intestinal message (3.623kb) and the somewhat larger message (∼3.8k) found in embryonic tissue at the gastrula and pluteus stages of development.
A TgHBox4 bacterial expressed protein that included the homeodomain was produced and antibodies were raised against it and were used in a collaborative study (Zhao et al., 1991) of the L1 late H2B histone gene in the sea urchin *Strongylocentrotus purpuratus*. Using gel shift assays, it was determined that the fusion protein specifically bound a radiolabeled oligo enhancer element derived from the L1 late H2B histone gene. In other gel shift assays, the antibodies were found to specifically block the binding of the TgHBox4 fusion protein as well as a specific enhancer binding nuclear factor to a radiolabeled oligo enhancer element like the one used in the previously described assay.

The antibodies raised against the TgHBox4 bacterially expressed protein were affinity purified (see Figure 6) and used to determine the localization of TgHBox4 proteins during embryogenesis (see Figure 7). The staining is generalized in the earlier stages with some localization where the primary mesenchyme cells position themselves on the blastocoel surface. In gastrula stage there is also some localization of TgHBox4 protein in the developing gut. By pluteus stage the TgHBox4 protein is localized to mesoderm cells of large intestine.

Speculation concerning the function of TgHBox4 proteins during embryogenesis can be made based on the binding studies carried out with the L1 late H2B gene and the immunolocalization studies discussed above, and knowledge of the Abdominal-B gene and it's homologs. Homeobox genes
code for proteins that include a sixty amino acid region that confers DNA binding ability. These genes have been found to regulate anterior-posterior patterning in arthropods and vertebrates. These genes are clustered and are regionally expressed in a collinear fashion with the regional expression of each gene being more anterior than the gene 5' of it. The Drosophila gene Abdominal-B is the most 5' gene the Antennapedia cluster, is expressed in the posterior parasegments 11 to 14, and is necessary for the correct development of the adult posterior segments 5 to 8. Although more numerous due to duplication events, the homologs of Abdominal-B in vertebrates display posterior expression patterns and are implicated in determination of cells in the region where they are expressed.

Because TgHBox4 is a homeobox gene that aligns with a gene in the HOM-C cluster, it is reasonable to expect it to contribute in the specification of the identity of a region during embryogenesis. Because TgHBox4 is an Abd-B homolog, it's effect during some point in embryogenesis is expected to be restricted to a posterior region of the developing embryo. Experimental evidence developed from my reagents implicates TgHBox4 as having a role in histone gene regulation. TgHBox4 proteins appear ubiquitously in the developing embryo from their appearance in blastulae until late blastula stage. During this stage of early embryogenesis, the TgHBox4 proteins maybe helping to maintain levels of the L1 late H2B gene product needed for expedient development of a rapidly growing embryo that consists of many cells that are in
the process of dividing. As the area of rapidly dividing cells becomes more limited, so would TgHBox4's expression pattern. After ectoderm cell division slows down, the level of cell division in the developing digestive system is still rather high. Expression of TgHBox4 in the digestive tract is to be expected then if one of its gene products is involved in L1 late H2B histone gene regulation.

It is also a possibility that low level generalized expression is necessary so that a TgHBox4 gene product is in every region. In this way the TgHBox4 gene product will be part of a unique set of transcription factors in a cell type and so contributes to the cell types ultimate determined state. This may result in a general cell type like ectoderm expressing a unique set of genes due to the unique set of transcription factors found within it. This idea is consistent with the changing expression pattern seen as the developing embryo gastrulates. A majority of TgHBox4's expression becomes limited to ectodermal cells where the primary mesenchyme cells position themselves after being released from the animal pole. It is very possible that TgHBox4 is involved in the control of expression of a gene coding for a cell adhesion protein. By regulating the expression of a cell adhesion protein, TgHBox4 may contribute to the determination of the pattern of the primary mesenchyme cells by locally altering (increasing) the affinity of certain ectoderm cells for emerging primary mesenchyme cells.
The endoskeleton of the developing embryo arises from organic material released from the primary mesenchyme cells. The significant element of the endoskeleton are large rods extending along the lateral margins of the posterior side of the developing pluteus. So it would not be out of character for a Abd-B homolog to be involved in primary mesenchyme cell arrangement, since a majority of the resultant structure is found in a posterior region.

As development proceeds, the areas of actively dividing cells becomes more limited and morphological regions become "permanently" defined. These two conditions lead to the prediction that TgHBox4’s expression pattern would become more limited to areas of actively dividing cells and/or to a posterior position where the gene's products would participate in defining the specific region's identity. The expression of TgHBox4 in the pluteus stage is consistent with this prediction. TgHBox4’s expression pattern becomes limited to the large intestinal region of the pluteus embryo. Here TgHBox4 may enhance the level of expression of the L1 late H2B histone gene in an area where cell division is occurring at levels higher than other tissues.

In general histone genes are expressed in all regions of a developing embryo but individual histone genes like the L1 late H2B gene may be regionally expressed. The TgHBox4 gene may be responsible for making sure adequate amounts of H2B histones are produced in rapidly dividing cells.

Alternatively, L1 late H2B gene production by TgHBox4 maybe part of a regulatory cascade leading to the identity of cell types in defined regions.
Histone proteins do interact with DNA and local differences in certain histone levels may result in altered chromatin structures. This could in turn affect the expression of other genes by making them accessible or inaccessible to transcription factors and DNA polymerases.

The analysis of the expression pattern of TgHBox4 during development leads to the postulate that TgHBox4 is at least a bifunctional gene. Northern analysis of the expression pattern of this gene during development and in adult tissues is consistent with this prediction. At blastula stage a major transcript of 4.4kb and minor transcript of 3.7kb is produced. By gastrula stage the 3.7kb message is being produced at levels equivalent to the 4.4kb message. By pluteus stage the 3.7kb message persists while the 4.4kb message is diminished to barely detectable levels. The protein from the 4.4kb message could have one or more functions. It may be involved in regulating the expression of an adhesion protein gene in the ectoderm where the primary mesenchyme cells arrange themselves. It may also function as an enhancer of the L1 late H2B histone gene. The product of the 3.7kb message that appears strongly in gastrula stage may continue to function in one of capacities discussed above. It may continue give an ectoderm region identity. Since the primary mesenchyme cell arrangement is already established by early gastrulation, when the 3.7kb message begins to appear, this smaller mRNA is unlikely to play a role at this location. Based on the expression pattern
detected by immunofluorescence, it is probably involved in the regulation of the L1 H2B histone gene in the still dividing developing digestive tract.

The expression of the TgHBox4 gene in adult tissue is consistent with the predictions stated above. A message of 3.6kb is found at significantly higher levels in adult tissues than the levels found at any time during embryogenesis. The tissues that TgHBox4 messages were discovered include large intestine, small intestine, ovaries, testis. These are all tissues that have a relatively high cell turnover rate during the life of the animal and so a high level of histone gene expression would be expected. High levels of TgHBox4 message is consistent with this expectation. The TgHBox4 message also appears in aristotle’s lantern. This is not a tissue of rapid turnover but is a tissue involved in mechanical movement. Cell replenishment would therefore be a necessity in times of overuse and histone gene expression would become important. Its constant high level of expression suggests other regulators are involved when cell division is necessary. Alternatively, TgHBox4 may be regulated in these cells at the post-transcriptional level.

The fact that the TgHBox4 messages found in adult tissues are smaller than any of the TgHBox4 messages found in the embryonic tissue is consistent with the above analysis. The sequences missing may have made the transcripts unstable and so lack of this region may result in stable more persistent messages. Unlike embryos which go through a series of developmental changes, adults morphology stay relatively constant. Instead of the need for
temporary expression of histone genes in cells that are temporarily rapidly dividing, their expression is needed constantly in tissues where cell turnover occurs throughout the life of the adult animal. This is consistent with the appearance of higher levels of the TgHBox4 message found in adult tissues that may or may not be more stable than message found in embryonic tissues.
APPENDIX A

Development of the sea urchin embryo

Spawning of male and female urchins in close proximity on the sea floor results in the beginning of the developmental process. A spermatozoan makes contact and enters the egg which produce an impulse around the perimeter of the egg. In response to this impulse cortical granules between the inner egg surface and the outer membrane (called the vitelline layer) swell and burst producing an outer ring encircling the egg called the fertilization membrane. It's main function seems to be blocking polyspermy fertilization.

Cell division begins about 90 minutes after fertilization with a cleavage furrow forming and the first cleavage occurring parallel to the egg axis resulting in two equal size cells (or blastomeres). The second and third division are also equal but the third cleavage is at right angles to the egg axis resulting in an upper and lower tier arrangement. During the next cleavage the top tier of cells divide equally and meridionally but the lower tier divide unequally into a set of larger upper cells and four smaller lower ones. The cells are now characterized from top to bottom as follows; a ring of eight medium size mesomeres, four large macromeres and four small micromeres. The fifth cleavage stage is characteristically asynchronous but ultimately results in a 32 cell stage embryo. The mesomeres divide into an upper and lower layer of cells. The mesomere "top" part of the embryo is the animal pole (an) and the cells from the top to the
bottom in this area are appropriately labeled an1 and an2. The micromeres divide unequally with smaller blastomeres being formed toward the "bottom" or vegetal pole of the developing embryo. At the sixth division the an1 and an2 cells divide horizontally resulting in 4 layers of 8 cells each. The macromeres also divide horizontally into an upper and lower vegetal(veg) layer of cells, called veg1 and veg2, respectively. The larger micromeres divide much later and the small micromeres don't divide at all until after later cleavages of cells of the upper part of the embryo. So technically cell stages can no longer be described in logarithmical terms because division has become so asynchronous at this point.

Early cell lineage studies of vitally stained blastoderm groups determined that an1, an2 and veg1 become ectoderm while veg2 become endoderm and secondary mesenchyme cells (Horstadius, 1935). More recent studies have determined the fate of these cells in even more specific terms. The embryo at this point is called a morula because of it's bumpy appearance resembling a mulberry. The cells now undergo a series of harder to follow divisions with the animal cells being much more active than the vegetal cells at this point. The cells become arranged in a circular pattern of single cells around a central cavity called a blastocoel and the embryo is now called a blastula. Each blastomere now develop a cilia and their beating result in the blastula rotating within the fertilization membrane. The embryo produces a hatching enzyme that dissolves the fertilization membrane and allows it to hatch to become a free
swimming organism. After hatching, the vegetal region of the blastula wall begins to thicken and the posterior wall of the blastula flattens where primary mesenchyme cells begin to separate and migrate into the blastocoel. At this point the embryo is made up of ~500 cells and is called a mesenchyme blastula (Davidson, 1986). Using thin extendable pseudopodia the mesenchyme cellsmove up the ventral wall about 1/4 of it’s total length and group on the right and left side. From here primary mesenchyme cells fan out in three directions, the center of which is the origin of the triradiate spicule which develops into the embryonic skeleton. Two of the sets of mesenchyme cells spread out along the ventral and dorsal walls and form a ring parallel to the posterior wall. The other set of cells go up the dorsal side to a distance of about 1/2 to 3/4 from the top. When all the primary mesenchyme cells are released the remaining cells of the posterior wall are primarily of veg2 origin and are presumptive endoderm cells (Horstadius, 1935).

The first step in gastrulation is when these cells move closer together and form what is called the endodermal plate. This plate starts to invaginate into the blastocoel cavity. Early in the process of the invagination of the endoderm which is now called the archenteron the secondary mesenchyme cells can be recognized at the top of the archenteron by their extending pseudopodia. These will extend up and attach near the animal pole on the ventral side of the embryo near where the mouth will eventually form. At this point the archenteron has become taller and thinner from the pull of the secondary
mesenchyme cells. Also after gastrulation the volume of the embryo, which has been constant for some time, starts to increase and pigment cells appear in the embryo for the first time.

As development proceeds the embryo changes shape. The ventral side of the embryo flattens and the dorsal side starts to bend in its direction. This curves the animal pole towards the ventral wall. The top of the animal pole rounds off at the ventral wall and is referred to as the oral lobe. At this stage the posterior wall spreads out also. After attachment of the archenteron to the ectoderm it begins to bend to meet the flat ventral wall. Here, descendants of an1 cells begin to invaginate forming a depression referred to as the stomodeum. The stomodeum and the archenteron both form openings where they meet to become the larval mouth.

While mouth formation is occurring the archenteron differentiates into three recognizable structures. The area nearest mouth formation narrows and becomes the esophagus, the middle expands to form the stomach, and the lower third of the archenteron becomes the intestine. The secondary mesenchyme cells at the top of the archenteron cluster into two aggregates which become coelomic vesicles, which are found at the sides of the esophagus.

While all this is taking place the spicules of the skeleton are also growing. First the two triradiate spicules form among the groups of mesenchyme cells in the right and left ventral corners. Like the primary mesenchyme cell
arrangement, one spicule radii points toward the animal pole and the other two of each triradial toward the dorsal and ventral sides. The spicules radiating out to the dorsal sides bend at a ninety degree angle to the plane of the spicular rudiment and grow down along the posterior wall to form structures called body rods. Where the bend starts a new spicule forms and grows in the opposite direction of the body rod. These structures are called the postoral rods. In *Tripnuestes gratilla* these rods actually consist of three rods originating from the spicule rudiment parallel to each other that are held together by cross bridges, forming what are termed fenestrated rods. The spicules originating from the radii facing the ventral side lengthen and become ventral transverse rods. Growth from the third spicule radii that face the animal pole grow for awhile to form the dorso-ventral connecting rods and then bend up towards the oral lobe to form the anterolateral rod.

As the ventral wall flattens the plane of the spicule rudiment changes somewhat. The affect of this can be seen the most on the posterior side. The origin of the body rods come closer together and the growth of the posterior side conforms to new arrangement. This results in the posterior wall taking the form of an isosceles triangle. Because of it’s novel shape this stage in the sea urchin's embryonic development is referred to as the prism stage.

The basic skeletal framework of the embryo has been laid down at this point though parts of it continue to grow as the embryo approaches it’s final stage of development. The body rods lengthen to the aboral side of the ectoderm
where their ends point towards each other and meet. The postoral rods continue to grow as the ectoderm around them lengthens to form the two long postoral arms. The anterolateral rods grow parallel to each other into the changing oral lobe. The oral lobe flattens dorsoventrally and widens. Ectoderm growth extends around the anterolateral rods to form the anterolateral arms, which are much shorter than the two postoral arms.

The embryo is now at its final stage before metamorphosis, a four armed free swimming pluteus. The digestive tract is now fully functional due to the completion of its development at this time including the esophagus becoming muscular and the development of cilia in the stomach. The embryo will remain a pluteus until the correct combination of time, temperature, and food will trigger it to begin metamorphosis. When this happens the coelomic vesicle on the left side of the esophagus will constrict to form a hydrocoel and the ectoderm wall adjacent to it will form a depression called the amniotic invagination. These two new structures together make up the echinus rudiment which is the imaginal disc in which the adult mouth, Aristotle's lantern and nerve ring are formed. Thus upon metamorphosis, the young adult will break out of the larval shell, only retaining the digestive tract of the embryo with the rest of the adult structures being formed after embryogenesis (Czhak, 1975).
APPENDIX B

Library of cTgHBox4-int cDNA sequencing reactions

This is a library of the results of sequencing reactions used to obtain the sequence of the first 1200 base pairs of the cTgHBox4-int (HB4CDNA), which includes all of the open reading frame. Each sequence is aligned with cDNA sequence that it represents. The template for sequencing reactions with T3, T7 or gene specific primers was either a clone with the complete 3.623kb cDNA, a clone with the 1.6kb 5’ EcoR1-EcoR1 fragment, or a clone that was a nested deletion (designated HB4SC for subclone) of the 1.6kb containing clone. Antisense sequences (3’ to 5’) have been inverted and complemented (designated by an I at the end of the name of each sequence).

<table>
<thead>
<tr>
<th>HB4CDNA</th>
<th>ATTGATAAATGGAGGAGATAACGTGACTGCTGCTGCCCTAAAAGGCCTG</th>
</tr>
</thead>
<tbody>
<tr>
<td>HB4INTT3A</td>
<td>ATTGATAAATGGAGGAGATAACGTGACTGCTGCTGCCCTAAAAGGCCTG 50</td>
</tr>
<tr>
<td>HB4SC63I</td>
<td>ATTGATAAATGGAGGAGATAACGTGACTGCTGCTGCCCTAAAAGGCCTG 50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HB4CDNA</th>
<th>TCCCAGCAGATTTACACTCTTTCTATCATTCATGGATCATCGAAATGCTCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>HB4INTT3A</td>
<td>TCCCAGCAGATTTACACTCTTTCTATCATTCATGGATCATCGAAATGCTCC 100</td>
</tr>
<tr>
<td>HB4SC525PI</td>
<td>-----------------------------------------CATGGATCATCGAAATGCTCC 21</td>
</tr>
<tr>
<td>HB4SC63I</td>
<td>TCCCAGCAGATTTACACTCTTTCTATCATTCATGGATCATCGAAATGCTCC 100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HB4CDNA</th>
<th>TTGACTCCGAAATTGCACGATCCGATCCAGGCCGCGCCGCGCGTCGTTG</th>
</tr>
</thead>
<tbody>
<tr>
<td>HB4INTT3A</td>
<td>TTGACTCCGAAATTGCACGATCCGATCCAGGCCGCGCCGCGCGTCGTTG 150</td>
</tr>
<tr>
<td>HB4SC525PI</td>
<td>---------------------------------------TTGACTCCGAAATTGCACGATCCGATCCAGGCCGCGCCGCGTCGTTG 150</td>
</tr>
<tr>
<td>HB4SC63I</td>
<td>TTGACTCCGAAATTGCACGATCCGATCCAGGCCGCGCCGCGTCGTTG 71</td>
</tr>
<tr>
<td>TTGACTCCGAAATTGCACGATCCGATCCAGGCCGCGCCGCGTCGTTG-- 127</td>
<td></td>
</tr>
</tbody>
</table>

136
HB4CDNA
GGATTTTACTCCATCGGTCGGCATCAAGCCTATGAACGATATCCTAATCA 450
92117PRIMI
GGATTTTACTCCATCGGATCAAGCCTATGAACGATATCCTAATCA 269
92275PRIME
GGATTTTACTCCATCGGATCAAGCCTATGAACGATATCCTAATCA 157

HB4CDNA
TGCTAACTCTATGTACCATCACTCTGCTGCAGCAGTGTCACACACGGGAC 500
92117PRIMI
TGCTAACTCTATGTACCATCACTCTGCTGCAGCAGTGTCACACACGGGAC 319
92275PRIME
TGCTAACTCTATGTACCATCACTCTGCTGCAGCAGTGTCACACACGGGAC 207

HB4CDNA
CCATACCTGTAAGCCTAACCGCTGCCGGGACTCTGGCGACGAAGTACATC 550
92117PRIMI
CCATACCTGTAACCGCTGCCGGGACTCTGGCGACGAAGTACATC 346
92275PRIME
CCATACCTGTAACCGCTGCCGGGACTCTGGCGACGAAGTACATC 239
HB4SC2F
CCATACCTGTAACCGCTGCCGGGACTCTGGCGACGAAGTACATC 51

HB4CDNA
ACCAGTGACGGTGGTGGTGGAAACACAAGCCCGCACGCTACAACACACAC 600
90204PRIMI
ACCAGTGACGGTGGTGGTGGAAACACAAGCCCGCACGCTACAACACACAC 9
HB4SC21LOI
ACCAGTGACGGTGGTGGTGGAAACACAAGCCCGCACGCTACAACACACAC 78
HB4SC2F
ACCAGTGACGGTGGTGGTGGAAACACAAGCCCGCACGCTACAACACACAC 101

HB4CDNA
CTTGCTTGACGAAAAGCACAAATACATCGCCAGTGGCGGTTATGAGGGAT 650
90204PRIMI
CTTGCTTGACGAAAAGCACAAATACATCGCCAGTGGCGGTTATGAGGGAT 59
HB4SC21LOI
CTTGCTTGACGAAAAGCACAAATACATCGCCAGTGGCGGTTATGAGGGAT 101
HB4SC2F
CTTGCTTGACGAAAAGCACAAATACATCGCCAGTGGCGGTTATGAGGGAT 151

138
HB4CDNA
ACAGCTGCCGAAAGAGCCCCTAGCGACTGTCAAATCGACGGGGCAGAGA 700
HB4SC1D
-------------------CCTAGCGACTGTCAAATCGACGGGGCAGAGA 31
HB4SC2F
ACAGCTGCCGAAAGAGCCCCTAGCGACTGTCAAATCGACGGGG------- 195
HB4CDNA
ACGGTGGGAGATCAGTCCTACGTACCCGTCGACGGGCTCGGCCGGTGGTTC 750
HB4SC1D
ACGGTGGGAGATCAGTCCTACGTACCCGTCGACGGGCTCGGCCGGTGGTTC 81
HB4CDNA
ACCTATTACAGTCATCCATAATAAGGACTCGCCTGAATCAGACATGAAAG 800
HB4SC1E
-----------ATTGATGACTGTGATGAGGACGAGAAGCAGAAGAACGGG 39
HB4SC1D
ACCTATTACAGTCATCCATAATAAGGACTCGCCTGAATCAGACATGAAAG 131
HB4CDNA
GACACGCCAACCTGGCTCTCGGCGACTTCCGGGAGAAAGAAGCGCTGTCC 900
HB4SC1E
-----------CTGGCTCTCGGCGACTTCCGGGAGAAAGAAGCGCTGTCC 39
HB4SC1D
GACACGCCAACCTGGCTCTCGGCGACTTCCGGGAGAAAGAAGCGCTGTCC 231
HB4CDNA
TTACACCAAGTTTCAGACGTTGGAGCTGGAGAAAGAGTTCCTCTTTAATA 950
HB4SC1E
TTACACCAAGTTTCAGACGTTGGAGCTGGAGAAAGAGTTCCTCTTTAATA 139
HB4SC1I
TTACACCAAGTTTCAGACGTTGGAGCTGGAGAAAGAGTTCCTCTTTAATA 89
HB4CDNA
TGTAACCTCACCCGCCAGACC8CGGCTTGAGATCGCTCTACTAAGCCTC 1000

HB4SC1E
TGTAACCTCACCCGCCAGACC8CGGCTTGAGATCGCTCTACAAAGCCTC 188

HB4SC12I
TGTAACCTCACCCGCCAGACC8CGGCTTGAGATCGCTCTACTAAGCCTC 139

HB4CDNA
ACCGAGCGGCAGTCAAAATCTGGTTCCAGAACGGAGAATGAAAGATGAA 1050

HB4SC1E
ACCGAGCGGCAGTCAAAATCTGGTTCCAGAACGGAGAATGAAAGATGAA 238

HB4SC12I
ACCGAGCGGCAGTCAAAATCTGGTTCCAGAACGGAGAATGAAAGATGAA 189

HB4CDNA
GAAGCAGAATCGGGCACAAAATTATCCCTAAATCCCACTGTAGTGCCAG 1100

HB4SC1E
GAAGCAGAATCGGGCACAAAATTATCCCTAAATCCCACTGTAGTGCCAG 288

HB4SC12I
GAAGCAGAATCGGGCACAAAATTATCCCTAAATCCCACTGTAGTGCCAG 239

HB4CDNA
TAATATAACTGTGACACCATGGAGCAGTTATCATTGGACATGAAGAAACA 1150

HB4SC1E
TAATATAACTGTGACCCATGGAGCAGTTATCATTGGACATGAAGAAACA 338

HB4SC12I
TAATATAACTGTGACACC8ATGGAGCAGTTATCATTGGACATG----- 279

HB4CDNA
CAATTGAATTGCATTTTGATGAATACTTGGAGATGACTCTAAGAGAAATG 1200

HB4SC1E
CAATTGAATTGCATTTTGATGAATACTTGGAGATGACTCTAAGAGAA--- 385

HB4SC1E
---------TTGGAGATGACTCTAAGAGAAATG 25
APPENDIX C

Complete cTgHBox4 DNA sequence

For DNA sequence HB4INTCDNA.
Analysis done on the complete sequence (3634) bases.

ATTGATAAAATGGAGAGAATAACGTGACTGCTGCTGCCCCTAAAAGGGCGCTGTCCACGAGA

TPTACACTCTTCTATATCTGAAATGTGACTCTTGACCCCTAATGGGAGACTGGCGAC

GACGCTCTACCTGACGACGCCGCCAATATATGCATACATAAACAGATTGACTGGCGAC

CTACGCATTCCTCTGTCAGCCGCCACCAACTCGGTACTGACTTTCCATGCTGCTG

CATACGTCCCCTTTTCTGAGGACTTACGACATGCTACCTTTAACCTACCATGCCGAT

GAAAAACTTTGAATCGGGACTAATGGCGGCTTCTGGAGGTGGATTTTACTCCATCGGTC

GCATCAAGCCTATAGATATCAGTAAACATCTGTAACTCTACACAGTACATCCATCTGCTC

AGCAGTGCACACAGGGACCCATACCTGTAAGCTACCCGCGCTGGGACCTGGGGAC

GAAGTACATCACCAGTGAGGGTGAGGGTGGAAACACAAGCCGCAGCTACACACACAC
<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTTGCTTGCAGAAGACAAATCAGCTGGCAGTGCGGTATGAGGGGATACAGCTGGCC</td>
<td>6</td>
</tr>
<tr>
<td>GAAAGACCCCCTAGCGACTGTCAATCGACGGGCCAGGGAAGACGGCCTACGAGAAAC</td>
<td>7</td>
</tr>
<tr>
<td>GTACCCTGTCGACGGGCTCGGCCGGTGGTTCACAATTACAGTCATCCATAAAAGGACTC</td>
<td>8</td>
</tr>
<tr>
<td>GCCCTGAAATCAGACATGAAAGAAATGACGTACATGTGACTGTGAGGACGAGAAGCA</td>
<td>9</td>
</tr>
<tr>
<td>MMB02</td>
<td>1</td>
</tr>
<tr>
<td>GAAGAACGGGACACGCACACGTGGCCAGACTTTCCGGGAGAAAAGGCGCTGTCC</td>
<td>2</td>
</tr>
<tr>
<td>TTACACAAAGTTTCAGACCTGGAGAAAGAGTTTCTCTTTAATATGTAACCTCAC</td>
<td>3</td>
</tr>
<tr>
<td>CCGAGACCACCACCCGCTTGAGATCGCTCTACTAAGCTCACCAGCGCCAGCTCAAAAAT</td>
<td>4</td>
</tr>
<tr>
<td>MMMB02</td>
<td>5</td>
</tr>
<tr>
<td>CTGGTCCAGAAACCGGAAATGAAAGATGAAAGAAACGAGAATCCGGGACACAAAAATTAC</td>
<td>6</td>
</tr>
<tr>
<td>AATCCCAACTGTAGTGCCAGTTATATCTAATGTGACACCATGGAGCAGTTATCATTGACA</td>
<td>7</td>
</tr>
<tr>
<td>MMB02</td>
<td>8</td>
</tr>
<tr>
<td>TGAAGAAACCAATGTGATTTTCATGAAACTTGCACATGAACTCTGCTATAGAATAAGAATG</td>
<td>9</td>
</tr>
<tr>
<td>CAGAAAACAGCAAGCGAGATGCTCCAAGTTTAATACATCGATATTAAAAGTTATGATTC</td>
<td>10</td>
</tr>
</tbody>
</table>
GTATTCATCATAACACAACATTTTAATGAGATTTATGAAACCCAAAACAGTAAATGTTCCGGTT
ACTTACATCGATTTTTAACCTTGAGCCATAATTATTTGCTTGCGTTGGAAGGTGGTTGGCCATATACGA
ACATAATCATTTTTAACAATAAAACGGTAAAGTTTTTTGCTCTAGGTGCGCCCTTAC
TGCGAGTTGAAAAATGACGACCAAGCAGTCTCAGACATGATTGAACATGCCGTGGATTATTGC
AGGACTGATATCAAGACAATGACCTCAGAACATGATTGAACATGCCGTGGATTATTGC
ATTTTGAAACGCTGATATGAAAGTATTCATTTTTCGTTTGCTTCCAATTTATTGTATAT
E 1
COR
TATATGCACATAAACAACCTTTGAAAAGTAAATGGAATTCACACAAAGGCTATTTCCTCC
CTTCAGATTATAAGATGTAGTAAATTGTTAGACCATATGCAATTTATCAACAGTCCTAT
CTCTTTCCTTTTTAATTCTTACCCCTCCACAGCATTCTCTCATTTGTAATTTAAACA
ATTGTTCACAAATATTTGATACTACATACACAGAAGTTTTAAAAAAGATCATTAAATATAC

143
AATAATACATGAATGGACACACACACACACATACACACATACCACGCACACACACA

CGCAACGCACACACACAACACATACACATACACACACACACACACACACACACACAC

TATTTTTTGAAGATTGTGGTATTATATGCTCAATAGAAAAACATACATCTACACACAA

ATAGTTTTTCAACGGCAAAACAAACTATTTAAAACATTACAGACCTATGTAATGTTAT

ATTGATAAATATTCACATCATAATACTACTCATATACAAATACGATTGATTTCAAAATGAA

TATGCTTTAAAAATGGTAAGAAATGTATCACTACTTCAGCATGTTAGTATTAATGTACA

CAAAAATCATTTTAAAGTAAAGTTGCGCCACCTTTGTAGACCTGTAGTATCTCGAG

ATACTATCTGAAAGCTACCCCGCTTACTTCTGGAATATCTTATCACAAAAGAAAGAA

ATAGAATAAAATATGCTAACATATCTTACATTGTCAGTTTTACGGCTATGTTAGA

CATTATTTAAAGAATATAACACACAAATACACATATATATATCTGATTTCAAAATACACCTA
ACTCTGAGTAAATTTGAAAAACCCAGAAACACAAACAAACAAACAAACAAACAAACAAATAGGGAATCG

ATGAGGAATAGTTAAACAAAAATATCTACCAGAAAAAGACCCTATCATAATGGTTACT

GGTTTTGATTAAATGAGATTTTAATTAAAGTAATAGGAGTTGTTCAAACAATATGATAT

TAATTTCACTATTGACTGTGACATTGACTTAATTGCATTGAATGTATTTGATCGACATT

GAAGCTGGTACAGCTTCATTCTTTTTCCACGAGAAACCGTGTCACATCGATTTACATAT

CAACTCTGCATTGCACTTTTTAAGACCAATATATCTTTAATCAAAACACACTTGG

CTATTATTGCTGGTAAATACATCTGAAAGCATTGCAATTCGATTAAATCAATGATT

ATATTCTGCGCTAAAACCTAATTCTGAAAGCATTGCAATTCGATTAAATCATGATT

TTTTGAATCCATTGACCTTTTTCCGTTAAAAGTTTTATGAAGCGTTAATATGGAGAATTG

TTTTAAGAGATATTTACTATAATTTAATTTGCTATC
BIBLIOGRAPHY


151


