Characterisation of *decapentaplegic* and other developmental genes in the cnidarian *Acropora millepora*

Gabrielle Natalie Samuel
Department of Molecular Biosciences
University of Adelaide

March 2002

Table of Contents

Abstract

Although many proteins essential for the correct development of animals have been well characterised in higher metazoans, the evolutionary roots of these molecules remain unknown. The phylum Cnidaria is the closest outgroup to the triploblastic higher metazoans and is likely to be critical to understanding the origins and evolution of higher metazoan development. This is a study of the molecular biology of the cnidarian *Acropora millepora*. Unlike the much studied cnidarian, *Hydra, A. millepora* is a member of the basal Anthozoan class of cnidarian, and should have characteristics that closely reflect the ancestral-state. Further, as a reef-building coral, *A. millepora* sexually reproduces during an annual mass spawning event, providing accessibility to embryonic material for study. Together, these facts make *A. millepora* an excellent system in which to analyse the origins of developmentally important pathways. In particular, members of the TGF-β-superfamily of signalling molecules are widespread in metazoans, but the evolutionary roots of its particular sub-classes are poorly defined. For example, it is well established that the DPP/BMP2/4 class plays a role in neural patterning, dorsal/ventral axis specification and limb development in both *D. melanogaster* and vertebrates. However, how and when these functions first arose is not known. This study provides evidence that *A. millepora* DPP is a functional homolog of *D. melanogaster melanogaster* DPP. Further, the presence of several components of a DPP/BMP2/4-specific signal transduction cascade are reported here, including a putative type I receptor and two putative receptor-activated Smads. Together these results indicate that DPP/BMP2/4 signalling predates limb development and raises the possibility that *dpp* plays a role in axis formation in simple non-bilateral animals. In order to facilitate the analysis of the origins of additional developmental pathways, this thesis also details a limited *A. millepora* EST study. This analysis generated potential *A. millepora* spatial and temporal marker genes, and a limited characterisation of one of these, *A. millepora hex,* is also reported here.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

I, Gabrielle Samuel, give consent to this copy of my thesis, when deposited in the University library, being made available for photocopying and loan.

Gabrielle Samuel

Acknowledgments

First and foremost I must of course acknowledge Rob who, over the past three years, has been there to offer advice, support and direction.

Thanks to Dave Miller for being such a great source of knowledge regarding all my thousands of evolutionary- and coral- related questions.

Thanks to everyone in the Saint lab, as well as the rest of the Genetics faculty. In particular, thanks to Sinead who, in the beginning was there to help with all the molecular work, to Tetyana for all the help with my fly work and to Donna, who was always readily offering help in just about every aspect of my project. A special thanks to Dan for all your help with those coral sequences, and of course for carrying all those nelly bins full of sea water while I lay in the sun!

Thanks to my collaborators in Canberra for all the support given to me on my visits, especially Dave and Nelida for the endless cups of coffee.

I must acknowledge the support I have received on a personal level from both my family and friends. For my friends, especially Rosa, this support came at times when, after every experiment had failed, they would partner me at the bar. For my family, the support was listening to me after I had been at the bar with my friends and was highly intoxicated!

To everyone I have failed to mention...a big thank you to you too!

Abbreviations

A_{600}: Absorbance at 600nm
Ac: acetate
AP: Alkaline phosphatase
APS: ammonium persulphate
BCIG: 5-Bromo-4-chloro-3-indolylβ-D-galactopyranoside
BCIP: 5-Bromo-4-chloro-3-indolyl phosphate
bis-acrylamide: N,N'-methylene-bisacrylamide
bp : base pairs
CIP:alkaline calf intestinal phosphatase
DAB: 3,3' Diaminobenzidine
DIG: Digoxygenin
DMF: dimethyl formamide
DNA: deoxyribonucleic acid
DNaseI: deoxyribonuclease I
dNTP: deoxyribonucleoside triphosphate
dH_2O: distilled water
DTT: dithiothreitol
EDTA: ethylenediaminetetraacetic acid
GST: glutathione-S-transferase
HRP: Horseradish peroxidase
IPTG: Isopropyl ß-D-thiogalactopyranoside
kb: kilobase pairs
kDa: kilo daltons
MQ: MilliQ
NBT: 4-nitro blue tetrazolium chloride
NP-40: non-idet P-40
NTP: nucleoside triphosphate
OD: optical density
PAGE: polyacrylamide gel electrophoresis
pfu: plaques
RNA: ribonucleic acid
rpm: revolutions per minute

RNaseA: ribonuclease A

RT: room temperature

SDS: sodium dodecyl sulphate

U: unit of enzyme activity

UV: ultraviolet light

V: Volts

v/v: volume per volume

w/v: weight per volume

1. Introduction

1.1 The study of evolution

In 1735 Linneus recognised that all the morphologically and physiologically distinct organisms could be divided and sub-divided into different groups and, based on morphological studies, he initiated a hierarchical classification system that ordered similar organisms together (see Figure 1.1A). Just over a century later, in 1859, Charles Darwin proposed an explanation for this classification system and thus became the pioneer of what has now become the very widely accepted theory of evolution by natural selection. The basis of this theory states that all organisms are related and are descendants of a single primitive common ancestor. In the Darwinian view, the history of life can be represented as a tree. The branches of the tree signify different species, the multiple branching and re-branching symbolising the dynamic progression of the evolution of life. All branches originate from a common trunk, which represents the single progenitor of life.

However, inferred evolutionary relationships between different organisms that are based on comparative morphology alone are inconclusive as many features can be wrongly interpreted. For example, false associations may result from convergent evolution. In recent years technology has increased to the point that organisms can be compared at the most fundamental level of their DNA, and it has become evident that evolutionary relationships are reflected in their DNA sequences. The application of molecular phylogenetics, which compares these sequences from various organisms, has provided a more precise classification system and has allowed the creation of an ever more accurate evolutionary tree.

From microscopic analysis it is evident that there is a division of life into unicellular and multicellular organisms. The transition from unicellularity to multicellularity, encompassing the specialisation of cells, has been placed approximately 2.1 billion years ago (reviewed by Campbell, 1993). Further, spatial differentiation, which allows the compartmentalisation of these specialised cell types, must have evolved in parallel (reviewed by Kirschner and Gerhart, 1998). This onset of complex organisation allowed new potential for the derivation of different species and resulted in the formation of a number of different kingdoms of multicellular organisms (reviewed by Campbell, 1993).

A.

Kingdom > Phylum > Class > Order > Family > Genus > Species

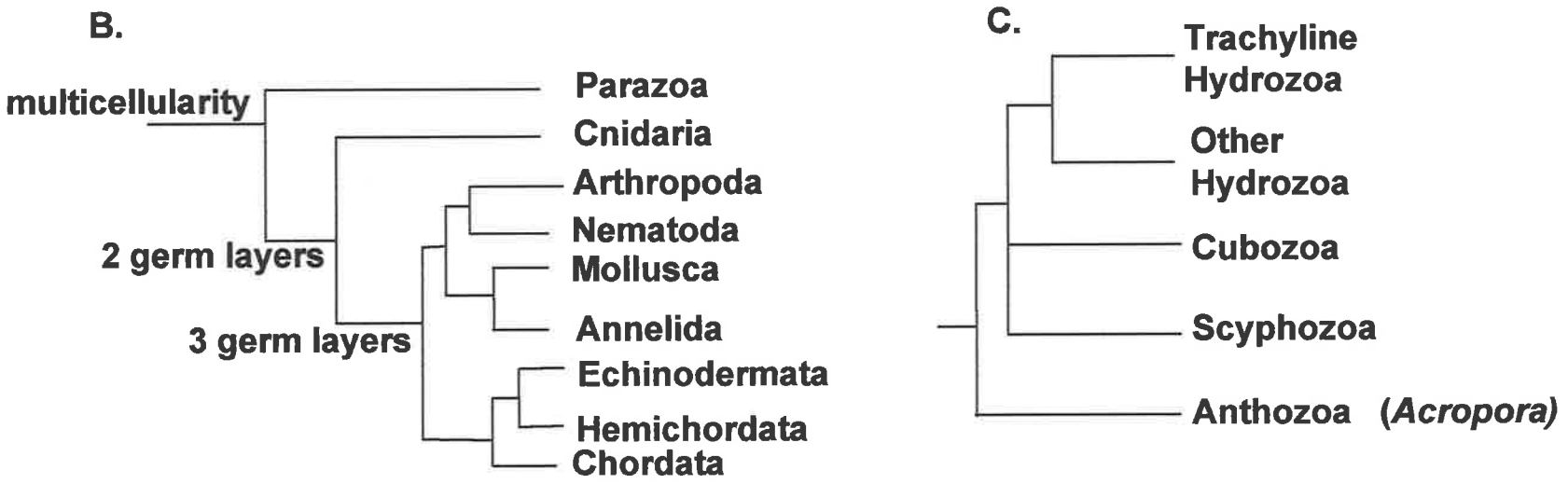


Figure 1.1. Classification of organisms and the evolutionary relationships of the metazoan phyla.
The major taxonomic groups of organisms are represented in A. B shows an evolutionary tree demonstrating the relationships between the major animal groups. C illustrates the class-level relationships within the phylum Cnidaria. 18S ribosomal DNA sequences and mitochondrial 16S ribosomal DNA sequences identify the Anthozoa as the basal class of cnidarian (Bridge *et al.*, 1992 and 1995). Taken from Miller and Ball, 2000.

1.2 The Metazoa

The Metazoa, or the kingdom of animals, can be defined as multicellular organisms that do not possess cell walls and pass through embryonic stages of development. The first metazoans appeared around 700 Mya (Million years ago), in the pre-Cambrian era, and their increased number of different cell types allowed an increased potential for the evolution of novel morphologies. However, evidence favours the presence of a less diverse collection of animals before a Cambrian explosion. This explosion, which dates back 540 Myr (Million years), led to a massive increase in diversity, producing a huge number of different species that encompassed a wide range of body plans (see Figure 1.1B; reviewed by Valentine *et al.*, 1994). The factors responsible for this sudden radiation of animals over 150 Myr after their initial appearance is still under debate, although the majority of theories encompass the change in atmospheric oxygen composition as a likely component (reviewed by Valentine *et al.*, 1999). Although the evolutionary origins of the Metazoa remained speculative for a long time, molecular phylogenetics has now confirmed that they are indeed monophyletic, all phyla originating from a single common ancestor (reviewed by Adoutte *et al.*, 1999; Holland, 1999).

Studying the development of animals is crucial to our understanding of evolutionary relationships between different phyla, as it lets us analyse how various body plans are patterned. Molecular phylogeny has allowed us to study and compare metazoan development at the molecular level, and has led to the discovery that the genes involved in directing body patterning during embryonic development are conserved over great evolutionary distances, similar genes being used for similar functions in very distinct phyla. This suggests that much of the basic gene regulatory machinery required to set up metazoan body plans was in place significantly before the Cambrian explosion (reviewed by Valentine *et al.*, 1999). The issue now is to discover what the differences and similarities are in conserved developmental systems, and how these systems have been re-deployed from a common metazoan ancestor to achieve independent morphologies in distinct phyla. In this way we can clarify which pathways are essential for multicellularity, and which evolved later to allow the subsequent diversification of the Metazoa. Solving these mysteries will be a major step in understanding the interplay of development and evolution, the aim of a new area of research interest, termed evolutionary developmental biology. This research is concerned with how developmental processes themselves have evolved: how they can be modified by genetic change, and how such modifications produce the past and present diversity of morphologies and body plans (Holland, 1999).

The functions of many developmental pathways have been well characterised through extensive studies of higher metazoan species such as *D. melanogaster, Caenorhabditis elegans* and vertebrates. These animals belong to distinct phyla, enabling a large amount of comparative data between animals that vary greatly in the organisation of their body plans to accumulate. However, simply documenting more cases of similarity between different higher metazoans will not help discover the evolutionary origins of developmental pathways. In addition, understanding the common principles of these pathways is complicated by the gene duplications and co-option of genes to the multiple developmental roles seen in higher animals. Because of this, the evolutionary roots of many of the broad classes of developmental pathways are unknown.

Analysis of the role of developmental pathways in primitive metazoans should provide valuable information regarding the origins of development and the evolution of complex body plans. The hope is that in these organisms fewer genes will play more restricted roles so that ancestral functions will be more clearly identifiable.

1.3 The Cnidaria

1.3.1 Cnidaria are primitive animals

With the exception of the sponges, the Cnidaria are thought to be the earliest evolved group of multicellular animals (see Figure 1.1B). Because of this they represent a major stage in the evolution of complexity and are potentially one of the most informative groups of organisms for studying the evolution of metazoan development. Further, hundreds of years of research on various species of *Hydra*, which is considered the textbook cnidarian, has provided a body of knowledge regarding the developmental and cell biology of this phylum. Four classes of cnidarians have been identified: the Hydrozoa, which includes the freshwater *Hydra vulgaris*; the Cubozoa and Scyphozoa, the jellyfish; and the Anthozoa, the corals and sea anemones. There is increasing evidence that recognises the Anthozoa as the basal class of the phylum Cnidaria, suggesting that animals in this class would be the most representative of ancestral characteristics (see Figure 1.1C; Bridge *et al.*, 1992 and 1995).

With few exceptions, cnidarians are marine animals. Unlike higher metazoans they retain a remarkable degree of plasticity at the cellular and tissue level. Thus, to a large extent tissue development is reversible, cells retaining the ability to de-differentiate and then re-differentiate into other forms (reviewed by Campbell, 1993). As primitive

animals, cnidarians are composed of an apparently small number of morphologically distinguishable cell types and have the simplest tissue structure; they have no true muscle, while nerves occur in their simplest form as a topologically two-dimensional nerve net (reviewed by Campbell, 1993). Cnidarians are diploblasts, developing through only a two germ-layer stage. This consists of an external ectoderm, giving rise principally to the epidermis and to nerve cells, and an internal endoderm, forming the digestive cavity. Diploblastic animals are essentially radial, possessing only one axis of symmetry known as the oral/aboral axis. The higher Metazoa are triploblastic, having a third germ layer, the mesoderm, lying between the other two, allowing the elaboration of muscles and internal organs. Triploblasts are inherently bilateral having two-sided symmetry: not only do they have a dorsal/ventral axis but they also have an anterior/posterior axis (reviewed by Campbell, 1993). The correspondence between the axis of radial animals and those of the bilateral ones is still under dispute.

1.3.2 Axis patterning in cnidarians

Recently there has been much discussion concerning the significance of the radial symmetry observed in cnidarians. Initially it was believed that bilateral animals descended from a more primitive radial ancestor. However, there are many examples within the phylum Cnidaria that show departures from true radial symmetry, with the only true radial cnidarian being *Hydra*. As recent evidence identifies *Hydra* as a member of the most divergent class of Cnidaria (Bridge *et al.*, 1992 and 1995) the existence of a bilateral metazoan ancestor has been suggested, with a trend to a loss of the second axis during cnidarian evolution. However, there is no evidence to support a simple correspondence between cnidarian bilaterality and the anterior/posterior and dorsal/ventral axes of higher metazoans. This implies that the weak bilaterality of cnidarians is likely to have arisen secondarily and independently of that seen in the bilateral animals.

An additional point of much interest concerns the origin of the axes of the bilateral animals. It was recently discovered that there are a group of genes, termed the Hox genes, which have been conserved throughout higher metazoans. The finding that these genes are clustered in the genome and that their chromosomal position corresponds to their expression domain along the anterior/posterior axis, initiated what is now known as the zootype theory (Slack *et al.*, 1993). This states that all animals have Hox gene clusters in their genomes, and that these genes have a conserved role in axis specification. When a number of Hox genes were initially identified as a cluster in the genome of certain cnidarians (for example, Schierwater *et al.*, 1991; Miller and Miles, 1993), it was believed that maybe the cnidarian single axis

was specified by Hox genes in a manner similar to the anterior/posterior axis of bilateral animals. However recent work suggests that this is incorrect, implying a possible role in axial patterning but not one necessarily consistent with the role of Hox genes in higher metazoans (Galliot, 2000; Gauchat *et al.*, 2000). Based on classical and morphological studies, others argue that the oral/aboral axis of cnidarians corresponds to the dorsal/ventral axis of bilateral animals (Willmer, 1990; Nielsen, 1995). The Wnt signalling cascade acts in the establishment of dorsal/ventral axes in higher metazoans. Thus, the recent discovery that it also functions in axis specification in the cnidarian *Hydra,* suggests a possible role for this signalling pathway in the axis specification of a common ancestor of both cnidarians and bilateral animals (Hobmayer *et al.*, 2000). However, again there is still no conclusive evidence of a simple correspondence between the dorsal/ventral axis and the oral/aboral axis. In addition, all these results leave open the possibility that signals for both anterior/posterior patterning and dorsal/ventral patterning existed in the radially symmetric metazoans, and became spatially uncoupled during bilateral evolution (Hobmayer *et al.*, 2000).

1.4 The coral *Acropora millepora*

1.4.1 *A. millepora* as a model cnidarian

The cell biology of the hydrozoan *Hydra* is well established because of the ability to analyse and interpret both gene function and expression patterns. However, *A. millepora* belongs to the anthozoan class of Cnidaria, which is likely to be more representative of ancestral characteristics than the much studied *Hydra*. Furthermore, as *Hydra* infrequently reproduce sexually, most studies of *Hydra* focus on its mechanisms of regeneration. Gene expression patterns during this process often differ radically from patterns seen during embryogenesis (reviewed by Gardiner *et al.*, 1995) and are therefore less informative when analysing embryonic development. In contrast, predictable annual mass spawning of coral sperm and eggs allows easy collection of *A. millepora* gametes. Fertilisation at a similar time point results in the synchronous development of embryos, facilitating the construction of developmentally distinct cDNA libraries. In addition, different to *Hydra*, library construction and transgenic work with *A. millepora* is not hampered by DNA instability due to an AT rich genome. An added advantage is seen when analysing *A. millepora* embryos. These are large compared to other cnidarians, aiding determination of the spatial distribution of gene products by *in situ* hybridisation or immunohistochemical techniques.

A. millepora belongs to the genus *Acropora,* which encompasses 368 species of reef-building hard corals that generally dominate Indo-Pacific reefs. The life cycle of *A. millepora* has been described to a limited degree (Miller and Ball, 2000) and consists of two distinct life stages, the sessile polyp and the planktonic embryo. The former is the dominant stage and the reef itself is composed of millions of individual polyps. The only contact a polyp has with the external medium is its mouth, which is the site of ingestion, removal of waste and gamete release. Protruding around the mouth are the tentacles, functioning in both defence and in the capture of prey. Polyps produce large colonies by asexual reproduction and all polyps in the colony remain connected to each other by extensions of their tissues. However, as described above, corals also reproduce sexually via an annual mass spawning of gametes, followed by external larval development (Harriott, 1983). The accessibility of *A. millepora* embryonic development makes it an excellent choice for research into the evolution of developmental pathways.

To date very little is known about *A. millepora* and its development. Representatives of several gene families have been isolated over recent years. These include Hox-related genes (Hayward *et al.*, 2001), a ubiquitin (Berghammer *et al.*, 1996), integrins (Brower *et al.*, 1997), nuclear receptors (Grasso *et al.*, 2001) and *ems* (David Miller, personal commun). However, no other genetic manipulations have been completed successfully. Analysis of gene expression patterns is still in the early stages of development and interpretation is complicated by the lack of both spatial and temporal marker genes.

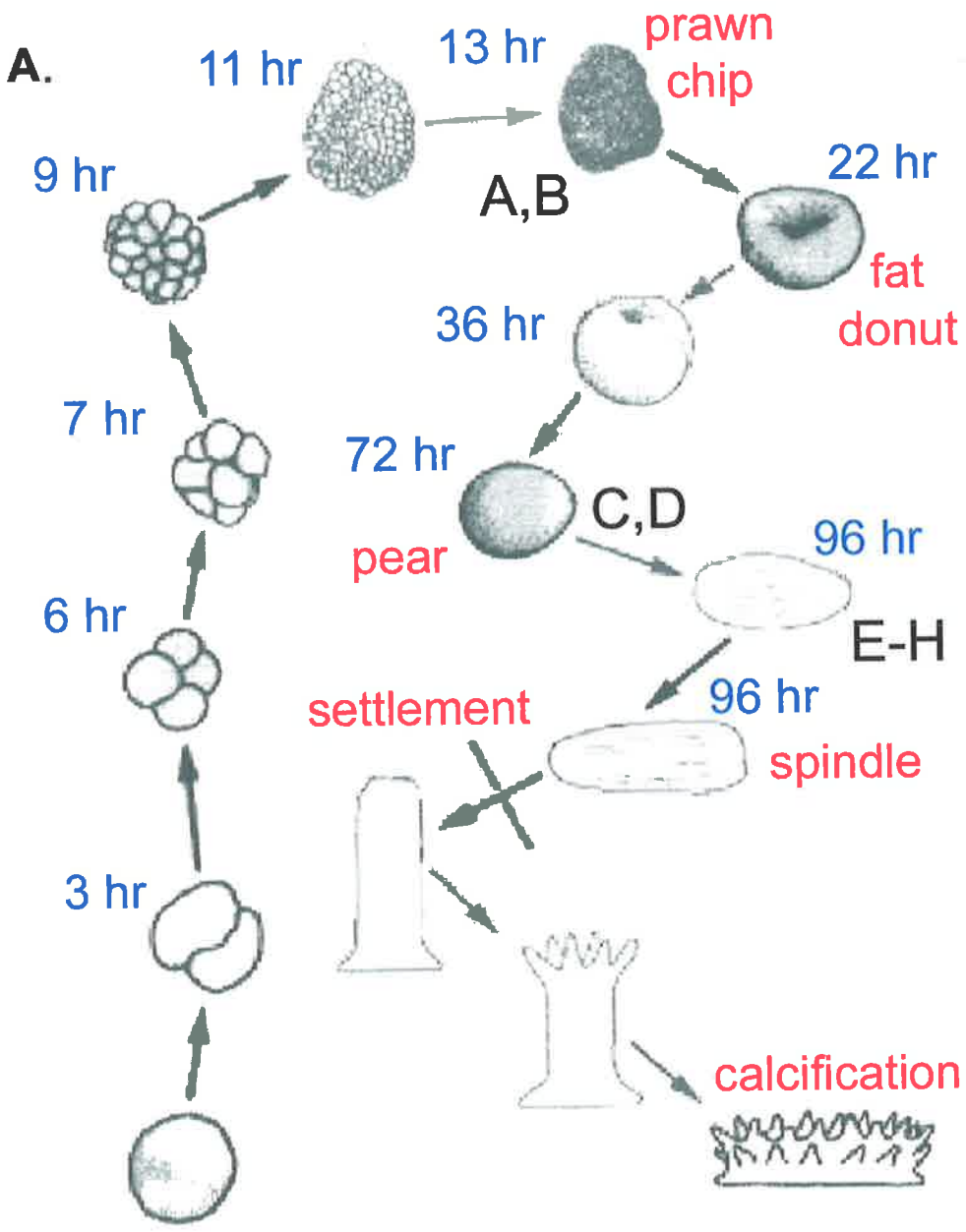
1.4.2 Development of *A. millepora*

After fertilisation, *A. millepora* undergoes radial cleavage and progresses through a number of easily identifiable developmental stages before settlement (reviewed by Miller and Ball, 2000; see Figure 1.2). Initial cell division leads to an irregular sphere of cells. This sphere continues to develop, and by 13 hours after fertilisation it forms a shape resembling a fried prawn chip, giving this stage its "prawn chip" name (see Figure 1.2; 600-800 μm). By a process not yet understood, gastrulation occurs and the two layers of cells comprising the embryo now separate with a depression developing on one side. This depression is believed to be the blastopore, and by now, approximately 22 hours after fertilisation, the embryo resembles a fat donut. This stage of development has thus been coined the "fat donut" stage (see Figure 1.2; 300-400 μm). As the blastopore closes, a second inner embryonic tissue layer, the endoderm, is formed. By approximately 26 hours after fertilisation the pore fully

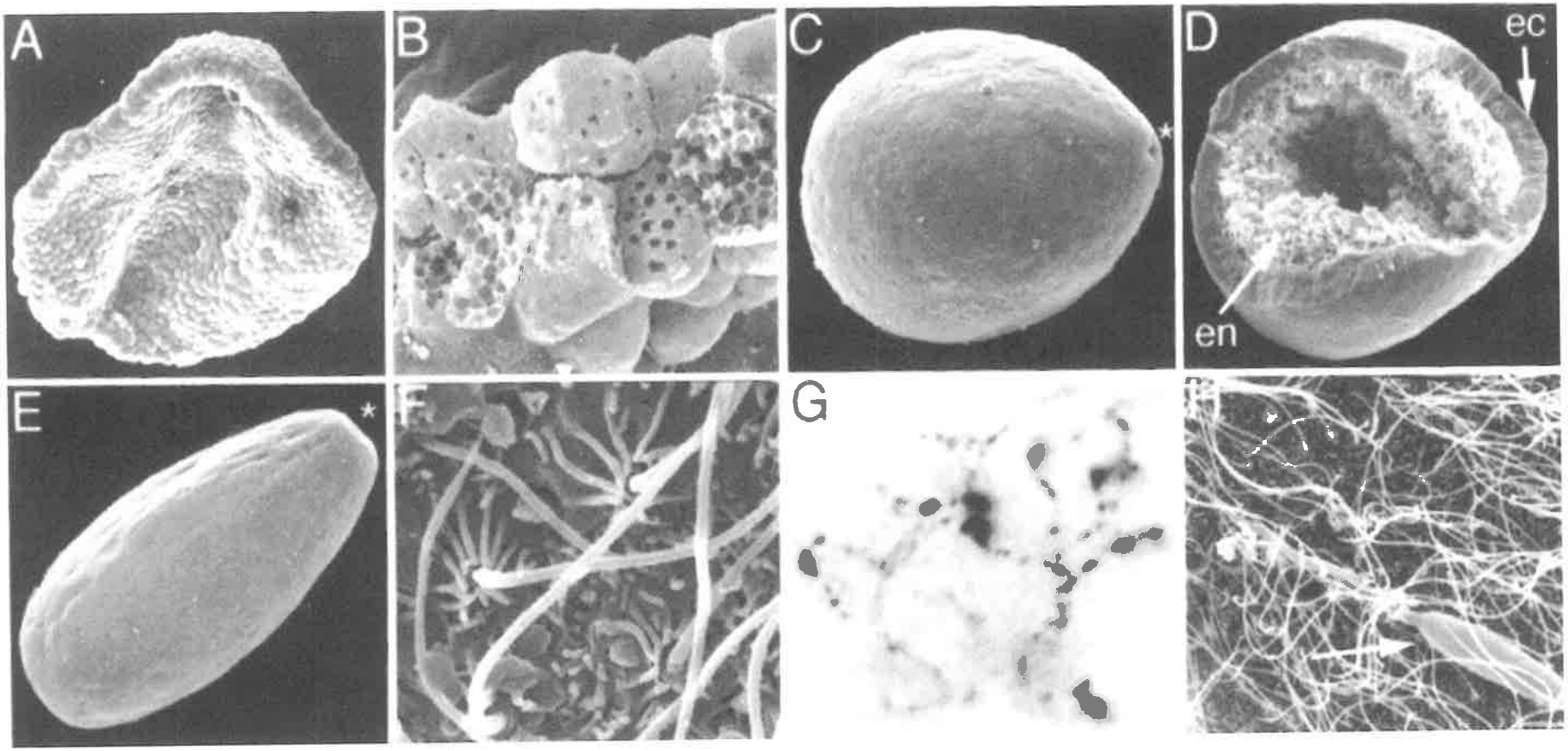
Figure 1.2. Embryonic development of *A. millepora.*

A: Schematic view of *A. millepora* embryonic development: Stages are not to scale. Times shown are representative and start at fertilisation, although they can vary considerably depending on temperature and other conditions. Large letters adjacent to a stage refer to the corresponding micrographic pictures in Figure 1.2B. Taken from Miller and Ball, 2000.

B: Micrographs of various embryonic stages of *A. millepora*: The prawn-chip stage is shown in A. It consists of a double layer of cells, as revealed from the cross-section shown in B. Later in development, at the pear stage (C, *= oral pore), the diploblastic nature of the embryo is clearly apparent in a cracked specimen (D, en= endoderm; ec= ectoderm). The fully developed planula (E, *= oral pore) is a sophisticated member of the zooplankton, with cilia for locomotion (F), a nerve net (G, RFamide staining) and nematocysts (H). Taken from Miller and Ball, 2000.


A.
11 hr
13 hr
prawn
chip
A,B
9 hr
22 hr
fat
donut
36 hr
7 hr
72 hr
C,D
pear
96 hr
E-H
6 hr
settlement
96 hr
spindle
3 hr
calcification

B.



closes, the embryo becomes more spherical, and external cilia appear (see Figure 1.2). This sphere becomes pear-shaped at around 72 hours after fertilisation, and an oral pore opens at the pointed tip of the pear (see Figure 1.2; 400-800 μm). This pear gradually elongates, forming a spindle shaped planula larva (700-800 μm). This is a cylindrical bilayer with external cilia that endow the planula with active swimming behaviour. The planula have a well developed nervous system, a mouth opening into the gastrovascular cavity, and nematocysts, which are presumably used for both defence and feeding. Prior to settlement the tissues de-differentiate into one cell type and then re-differentiate into the adult coral (for example, Vandermeulen, 1974 and 1975). Settlement involves attachment to the substratum and metamorphosis into a juvenile polyp (Miller and Harrison, 1990; reviewed by Miller and Ball, 2000)

1.5 Intercellular signalling pathways

Eukaryotic organisation and development of correct body plans is dependent on the communication between different cell types which, in the Metazoa, is provided by intercellular signalling factors. As metazoans evolved so did ever more complex spatial organisations of multiple specialised cell types, and it therefore seems probable that one key factor important in the increasing complexity of the Metazoa is the corresponding increase in the diversity of intercellular signalling systems. This is most evident in vertebrates, which use many more signalling molecules than do invertebrates. Because most or all of these broad classes of systems are also present in *D. melanogaster* and *C. elegans* we can understand this complexity in vertebrates in terms of genome wide duplications that are likely to have occurred in the chordate lineage (Holland, 1998). Studying intercellular signalling pathways and their functions during development of more primitive organisms is a vital step in understanding how the complexity of the animal kingdom arose.

As primitive metazoans with few cell types, cnidarians might be expected to have only limited cellular signalling requirements, perhaps a few more than the sponges, but many fewer than the triploblastic animals. To date there is clear evidence in both sponges and cnidarians for the presence of G-protein coupled receptors (Vibede *et al.*, 1998) and integrins (Brower *et al.*, 1997). Further, the presence of a probable FGF receptor (Genbank #AF070966) implies that cnidarians are likely to make use of transmembrane tyrosine kinase signalling pathways. In addition, the Cnidaria have various types of septate (gap) junctions (eg. Green and Flower, 1980) and key components of nuclear hormone receptor-mediated (Kostoruch *et al.*, 1998) and Wnt (Hobmeyer, 1996) signalling pathways.

Despite the fact that the serine/threonine (Ser/Thr) signalling pathway has been well characterised in vertebrates, *Drosophila* and *C. elegans*, this very important class of signalling mechanisms is poorly studied in diploblasts. With the current lack of comparative data concerning this signalling system in a primitive metazoan, evolutionary information may be gained from the analysis of this pathway in the cnidarian *A. millepora*. For this reason it was decided to explore one aspect of transmembrane Ser/Thr signalling, that represented by Decapentaplegic (DPP)/ Bone morphogenetic protein (BMP) 2/4 members of the TGF-β superfamily of signalling molecules. Below, the TGF-β superfamily is described briefly, as an introduction to a more detailed description of the DPP/BMP2/4 sub-family.

1.6 The TGF-β superfamily of intercellular signalling molecules

1.6.1 The TGF-β family

The TGF-β superfamily of signalling molecules encompasses a diverse range of multifunctional proteins, which have been identified throughout both vertebrates and invertebrates. Functions of these proteins include the control of proliferation, apoptosis, morphogenesis and specification of cell fate (reviewed by Massagué, 1998). Based on sequence alignment between members of the TGF-β superfamily, several sub-families have been identified. The *D. melanogaster* DPP protein and its vertebrate homologs BMP2 and 4 form one sub-group. Other groups include the TGF-β and activin sub-families.

1.6.2 The TGF-β ligands

Members of the TGF-β superfamily are initially synthesised as precursor molecules. These precursors can be divided into three main regions (see Figure 1.3). At the N-terminus is a hydrophobic signal sequence. This sequence, indicative of transmembrane and secreted proteins, is involved in secretion of the ligand. Adjacent to this signal peptide is the proregion. The sequence of this region is poorly conserved between family members, but may be conserved in any particular member between species. The 110-130 C-terminal residues will eventually form the biologically active protein, and it is the amino acid sequence conservation of these residues that characterise family members. The cleavage site that releases the active protein, located between the proregion and the C-terminal domain, is normally a dibasic or RXXR (R = arginine; X = amino acid residue) site (reviewed by Massagué, 1990 and Kingsley, 1994). After cleavage, and prior to secretion, the ligands dimerise. In addition to homodimers, heterodimers can form in cells expressing two different

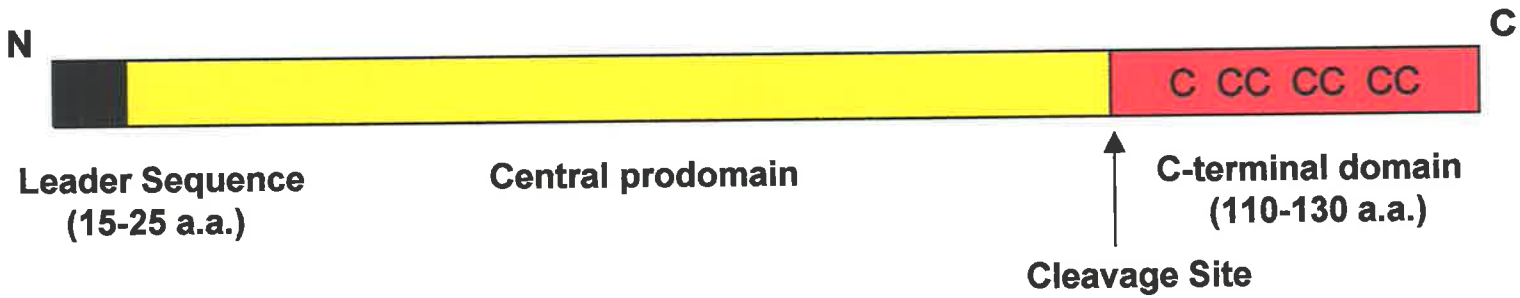


Figure 1.3. Schematic of a TGF-β superfamily ligand.

There are three main regions: the leader sequence, the prodomain and the C-terminal domain. The C-terminal domain is cleaved to form the active protein at a cleavage site, as indicated by the arrow. C = conserved Cys residue in the active protein; a.a. = amino acids; N = N terminus; C = C terminus. Not to scale. Taken from Kingsley, 1994.

TGF-β family members. Dimerisation occurs, at least in part, via disulphide bond formation (reviewed by Massagué, 1990), presumably involving the 7 conserved cysteine residues in the N-terminal region of the active ligand (reviewed by Kingsley, 1994).

1.6.3 The TGF-β transmembrane proteins

1.6.3.1 The receptor serine/threonine kinases

Initial speculation that members of the TGF-β superfamily of ligands signal through a distinct family of receptors arose after the identification of the activin receptor, ActIIR (Matthews and Vale, 1991). This receptor had a kinase domain that showed 32% similarity to the kinase domain of the *C. elegans* DAF-1 receptor (Georgi *et al.*, 1990), yet little similarity to other known receptor kinase families. Later, expression cloning of the human TGF-β II receptor (Lin *et al.*, 1992) revealed that its kinase domain too, showed similarity to that same region of DAF-1 and ActIIR. The conservation of sequence within this domain allowed identification of additional family members via PCR methods, revealing a number of similar features shared between the proteins. This novel family of receptors was termed the receptor serine/threonine (Ser/Thr) kinases.

1.6.3.2 Receptor serine/threonine kinase structure

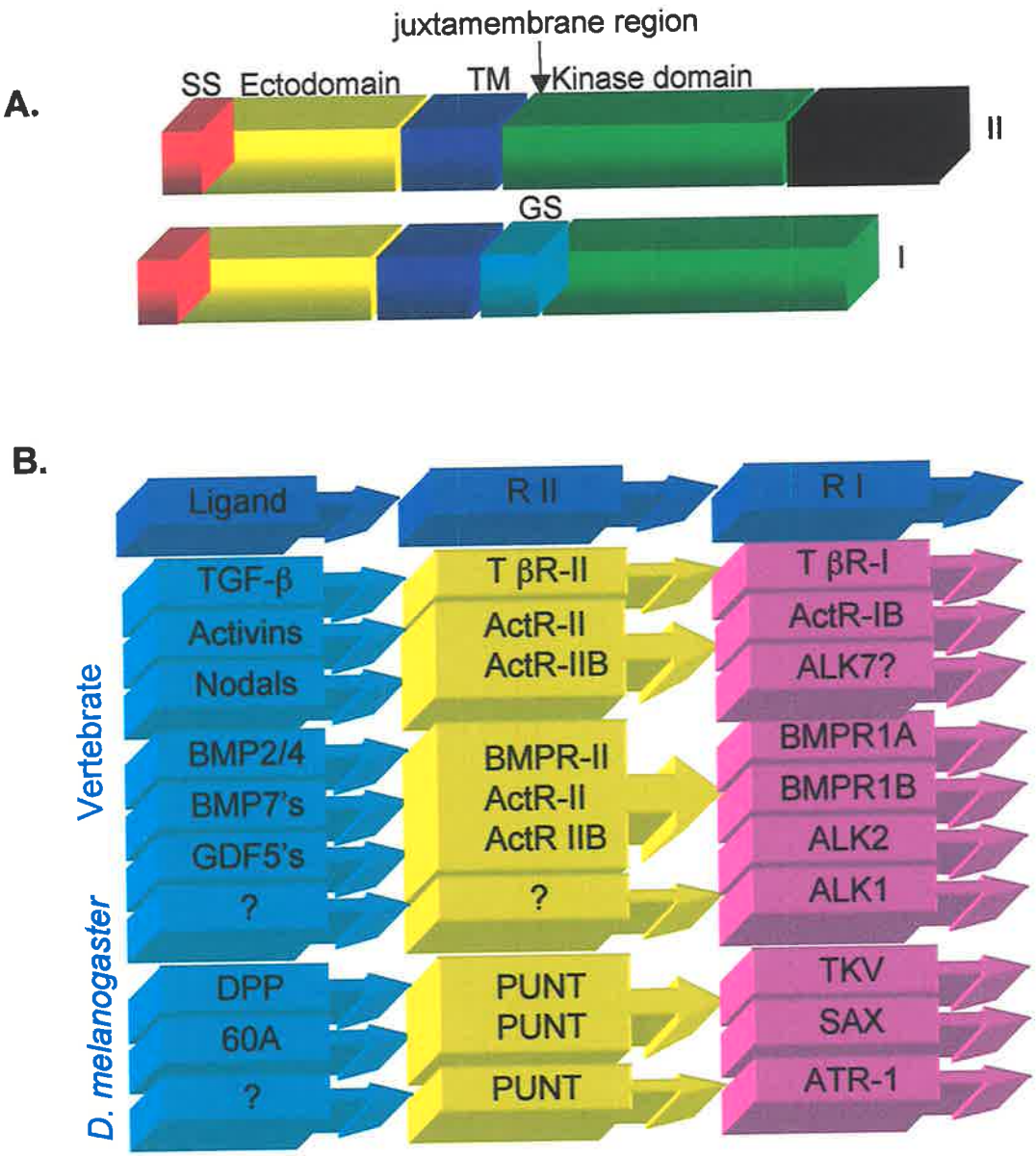
The N-termini of receptor Ser/Thr kinases are extracellular. Similar to nearly all membrane and secreted proteins, the initial 20-30 amino acid residues form a short hydrophobic signal sequence that directs the protein to the cytoplasmic membrane. The remainder of the extracellular region is generally between 100 and 170 amino acid residues in length and contains a 9 amino acid cysteine box, $CCX_{4-5}CN$, which suggests the potential for protein-protein interactions occurring via di-sulphide bond formation. It is this region that is capable of responding to a specific ligand. A single transmembrane domain passes through the membrane in an α-helical structure, followed by the intracellular C-terminal region. This latter region incorporates a kinase domain, typified by 12 sub-domains containing a number of highly conserved amino acid residues. Sequences in two of these domains, VIB and VIII, determine Ser/Thr kinase specificity (reviewed by Massagué, 1998). More detailed analysis of receptor Ser/Thr kinases illustrates that they exist as 2 distinct types of receptor, type I and type II. Although they have a similar overall composition, the two types differ in structural detail, sequence similarities, and function.

Type II receptors: These receptors are 70-80 kDa. The extracellular domain varies between 130-170 amino acid residues. In addition to the cysteine box there are a

Figure 1.4. TGF-β receptors and their interactions during signalling.

A: Schematic of a TGF-β type I and type II receptor: Receptor type I is shown above receptor type II. Differences between receptors include an extended Ser/Thr tail in the type II receptor and a GS rich region in receptor type I. SS = signal sequence; TM = transmembrane domain; GS = glycine/serine rich domain. The juxtamembrane region is indicated. Not drawn to scale.

B: TGF-β receptor interactions during signalling: Schematic showing specific ligand-receptor II and receptor II-receptor I interactions that occur during TGF-β superfamily signalling in both vertebrates and *D. melanogaster*. The columns, from left to right, list the ligands, the type II receptor and the type I receptor, respectively. Adapted from Massagué and Wotton, 2000.


A.
juxtamembrane region
SS
Ectodomain
TM
Kinase domain
II
GS
I
B.
Ligand
R II
R I
TGF-β
Activins
Nodals
T βR-II
ActR-II
ActR-IIB
T βR-I
ActR-IB
ALK7?
BMP2/4
BMP7's
GDF5's
?
BMPR-II
ActR-II
ActR IIB
?
BMPR1A
BMPR1B
ALK2
ALK1
Vertebrate
DPP
60A
?
PUNT
PUNT
PUNT
TKV
SAX
ATR-1
D. melanogaster

further 10 cysteine residues which are variably positioned in the extracellular domain. The kinase domain is typically 290-300 amino acid residues, and is followed by an extended C-terminal cytoplasmic tail which is variable in length, generally between 24-90 amino acid residues, and Ser/Thr rich (see Figure 1.4A; reviewed by Massagué, 1992; Kingsley, 1994).

Type I receptors: These are approximately 55-60 kDa. They too have additional extracellular cysteine residues that precede the cysteine box, but the positions of the cysteines are more conserved than those of the type II receptors, and have the sequence [$CXCX_{3-5}CX_{4-25}CX_{4-5}CX_{13-16}GCX_{7-19}CX_{11-13}$]. Type I receptors have no Ser/Thr rich C-terminal extensions. Instead they have a glycine/serine rich sequence, SGSGSG, termed the GS domain, situated adjacent to the juxtamembrane region of the intracellular domain (see Figure 1.4A; reviewed by Massagué, 1992; Kingsley 1994).

Phylogenetic analysis demonstrates that both type I and type II receptors form a single large lineage from a common ancestor. This tree does not bifurcate into two monophyletic groups representing each type of receptor, but reveals that type I receptors repeatedly diverged from a type II receptor (Newfeld *et al.*, 1999). The data reveals more sequence divergence in the extracellular domain than in the kinase domain. In addition, it shows a number of receptor Ser/Thr kinase sub-families. Phylogenetic analysis using the extracellular domain grouped receptors into different sub-families than phylogenetic analysis conducted on the kinase domains. However, using either domain, type I and type II receptors never grouped together into the same sub-family (Newfeld *et al.*, 1999).

1.6.3.3 Receptor-ligand interactions

Cell culture studies testing interactions between different receptors and their ligands have demonstrated the formation of a heteromeric complex between certain type I and type II receptors, as well as specific ligand interactions with these complexes (for example, Wrana *et al.*, 1992; Attisano *et al.*, 1993; Ebner *et al.*, 1993; Franzen *et al.*, 1993; ten Dijke *et al.*, 1994). There are two modes of ligand binding. Type I TGF-β/activin receptors recognise their ligand bound to receptor II but do not recognise the free ligand (Wrana *et al.*, 1992). Ligand binding therefore seems to be determined by receptor II, which specifically complexes with the ligand in the absence of receptor I allowing each type I receptor to bind and interact with a number of ligand-receptor II complexes. In contrast, BMP ligands bind both receptor types independently, yet co-operatively, where they jointly form a high affinity

complex (Liu *et al.*, 1995; Nohno *et al.*, 1995; Rosenzweig *et al.*, 1995; Nishitoh *et al.*, 1996). It is thought that the type II receptor pre-exists as a homodimer *in vivo*, as the presence of a very stable homodimer *in vitro* has been documented, such that this homodimer can still associate in the presence of strong detergents (Chen and Derynck, 1994). As a homodimer this receptor provides two type I receptor binding sites, predicting the formation of a receptor tetracomplex containing two ligand dimer binding domains (Yamashita *et al.*, 1994). Figure 1.4B summarises the various interactions that can occur between a ligand, and the two receptor types. This sharing of ligands between different receptors may potentially allow cross-talk between all the members of the TGF-β superfamily.

Yeast-two-hybrid systems have been employed to look more closely at the specific interactions between ligands and receptors. In particular they have demonstrated that the cytoplasmic domains of receptor type I and type II have intrinsic affinities for each other, and are therefore likely to play a role in both the formation and stabilisation of heterocomplexes (Liu *et al.*, 1995). These regions, however, are not an absolute requirement for complex formation, as truncated receptors lacking cytoplasmic domains can still associate (Chen *et al.*, 1995). This suggests that there are multiple contact sites along the receptors (Chen *et al.*, 1995).

1.6.3.4 Receptor signalling

Analysis of various chimeric receptors has provided evidence that the intracellular domains of each receptor have a distinct, non-interchangeable, functional role in the heterocomplex, and that both intracellular domains are necessary for signalling (Muramatsu *et al.*, 1997). Further experiments, which analysed the phosphorylation states of receptors TRβ-I and -II, showed this signalling to be transduced by phosphorylation events: The type II receptor is highly phosphorylated in the absence of a ligand. Addition of a ligand does not alter the overall phosphorylation level demonstrating that phosphorylation is constitutive (Wrana *et al.*, 1994). Similar experiments demonstrated that the type I receptor is only phosphorylated when in a heterocomplex with the receptor type II, and that this ligand-induced phosphorylation of receptor type I is catalysed by the kinase activity of receptor type II (Wrana *et al.*, 1994). Phosphorylation of the type I receptor was shown to occur in the GS domain (Wrana *et al.*, 1994). Furthermore, receptor complexes in which the GS domain is not phosphorylated are unable to transmit signals, suggesting that the GS domain is a key regulatory region that may control catalytic activity of the type I kinase or interaction with a substrate (Wrana *et al.*, 1994).

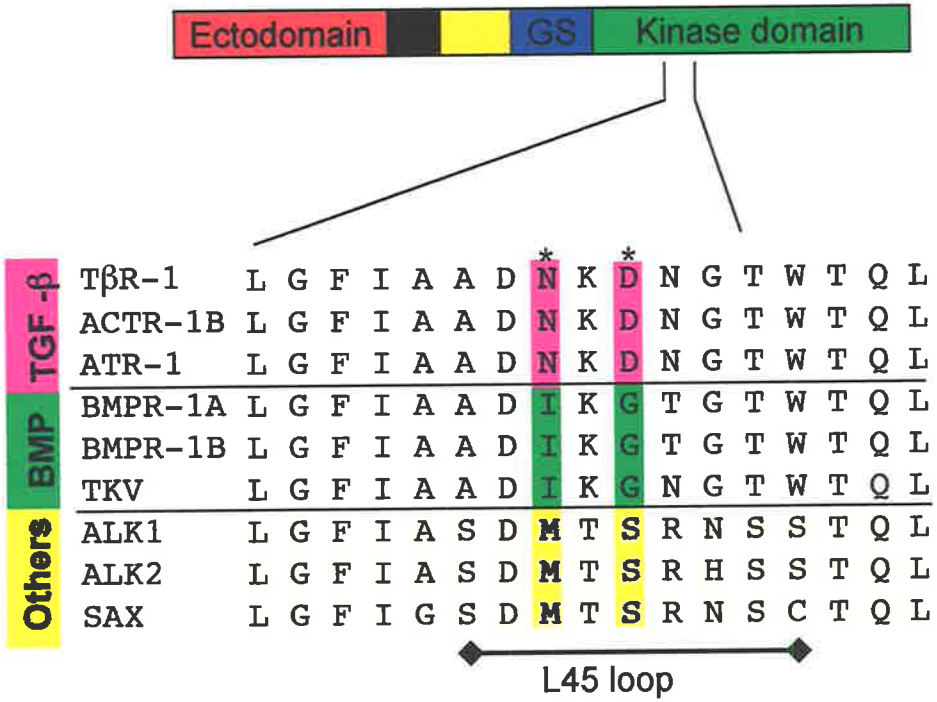


Figure 1.5. Alignment of the L45 sequences from various TGF-β type I receptors.
The L45 loop is present in the intercellular region of type I receptors. It is responsible for sub-class-specific signalling. The residues constituting the L45 loop are marked at the base of the figure. The specific TGF-β receptor sub-class is shown on the left. Note the two sub-type specific amino residues represented by an *. Taken from Chen *et al.*, 1998.

1.6.3.5 Receptor Specificity

The majority of the kinase domain and the juxtamembrane region of different type I receptors are functionally equivalent. However there is an area, designated the L45 loop, which has been shown to confer specificity of the receptor for downstream components of the pathway (Feng and Derynck, 1997; Chen *et al.*, 1998). In TβR-I, this region lies between the amino acid residues 263 and 271. The sequence of this region is conserved in receptors from similar TGF-β family sub-groups, but differs between receptors from different sub-families (see Figure 1.5).

1.6.4 Intracellular transduction of the TGF-β signal

1.6.4.1 The Smad proteins

The initial Smad protein was identified in a screen for dominant enhancers of the *dpp* phenotype in *D. melanogaster* imaginal discs. The protein was subsequently named Mothers Against DPP (MAD) and was shown to have a mutant phenotype similar to that of *dpp* in all aspects of *dpp* function (Sekelsky *et al.*, 1995). Comparison of the MAD protein sequence with other sequences revealed similarity with three putative *C. elegans* genes. These genes were later confirmed to encode proteins involved in the *C. elegans* TGF-β signalling pathway and were termed *sma-2, -3* and *-4;* all functionally related but not redundant (Savage *et al.*, 1996). The respective proteins were subsequently shown to have an intracellular function (Savage *et al.*, 1996) and to act downstream of the receptor Ser/Thr kinases (Thomsen, 1996).

Degenerate PCR, further genetic screens and analysis of ESTs in other vertebrates and invertebrates, identified additional Smad homologs. All encode 400 amino acid residue proteins and have two very highly conserved domains - the MH1 domain in the C-terminal half, and the MH2 domain in the N-terminal half. These are linked by a poorly conserved proline-rich linker region (Heldin *et al.*, 1997; reviewed by Massagué, 1998). To date nine different vertebrate Smads have been identified and these proteins can be classified into three Smad classes based on both sequence and function criteria (see below). Figure 1.6A illustrates each of the Smads, classifying them into their respective sub-families.

1.6.4.2 Receptor-regulated (R-) Smads

R-Smads are the primary targets of the activated Ser/Thr kinase receptors. Different R-Smads are activated by the binding of specific TGF-β superfamily ligands to their receptors: In experiments where various R-Smads were injected into *Xenopus laevis* animal pole explants (caps), the effects of Smad1 and the homologous Smad5 and Smad8 mimicked those of BMP2/4 ligand injection. This suggested that they primarily

act in the BMP2/4 signalling pathway (Baker & Harland, 1996; Graff *et al.*, 1996; Liu *et al.*, 1996; Thomsen, 1996). Injection of Smad2 and Smad3 induced effects similar to that of TGF-β and activin injection, demonstrating their involvement in the activin/TGF-β pathway (Graff *et al.*, 1996; Zhang *et al.*, 1996). Further studies of R-Smad-receptor interactions suggested that these R-Smads transiently associate with the receptor complex. This association is via an interaction with receptor type I whereby Smad1, 5 and 8 associate with type I receptors involved in BMP2/4 signalling, and Smad2 and 3 interact with activin/TGF-β-specific receptors (Macias-Silva *et al.*, 1996; Nakao *et al.*, 1997a).

Mutational analysis demonstrated that an L3 loop present in R-Smads is both essential and provides specificity for the interaction with receptor type I (Lo *et al.*, 1998). The loop is a highly conserved region in the MH2 domain that, by analogy to the Smad4 (see later) crystal structure, protrudes out from the molecule poised for protein-protein interactions (Shi *et al.*, 1997). The sequence of this loop is invariant bar two amino acid positions, which are sub-type-specific. Residues in these positions are identical in TGF-β-activated R-Smads, and in BMP-activated R-Smads, yet different between these two groups (see Figure 1.6B). Evidence has been provided for direct interaction between this L3 loop and specific L45 regions of the type I receptors (Chen *et al.,* 1998).

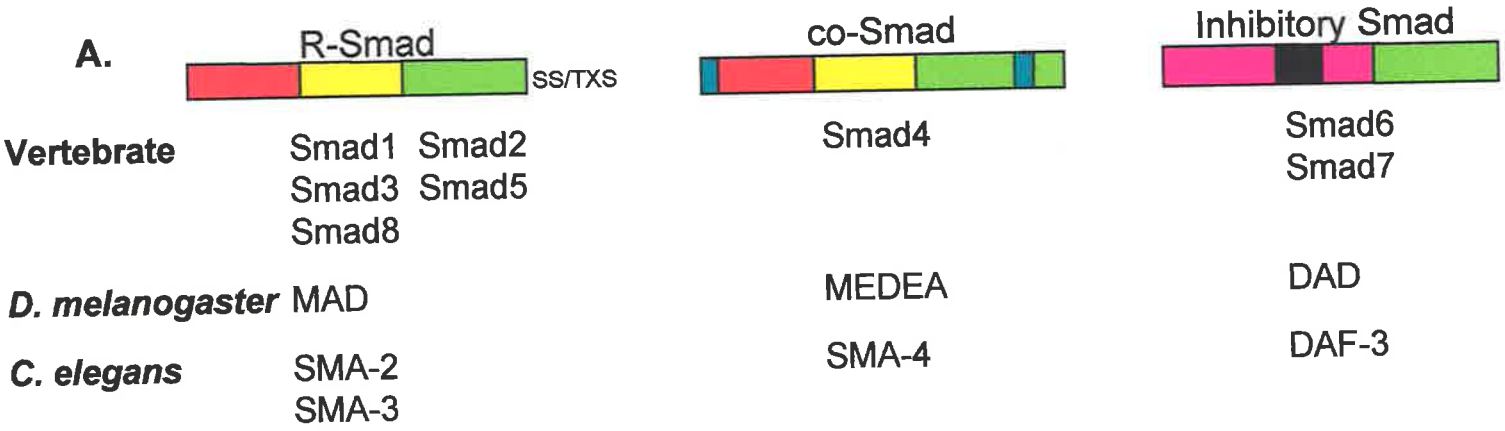
Examination of phosphorylation states of the R-Smads indicates that they are phosphorylated in response to their specific ligand (Eppert *et al.*, 1996; Hoodless *et al.*, 1996; Nakao *et al.*, 1997a). Further, *in vitro* kinase assays demonstrate that this phosphorylation is induced directly by the type I receptors (Macias-Silva *et al.*, 1996; Kretzschmar *et al.*, 1997). Ligand induced phosphorylation occurs at an SS/TXS site present at the C-terminus of the R-Smads (Macias-Silva *et al.*, 1996; Kretzchmer *et al.*, 1997; Souchelnytskyi *et al.*, 1997). Phospho-amino acid analysis demonstrates that phosphorylation of these R-Smads occurs mainly on serines with few threonines and no tyrosines being phosphorylated (Hoodless *et al.*, 1996; Lechleider *et al.*, 1996; Nakao *et al.*, 1997b). Blocking phosphorylation prevents signal-dependent activities, illustrating the importance of R-Smad phosphorylation in signal transduction (Macias-Silva *et al.*, 1996; Wisotzkey *et al.*, 1998).

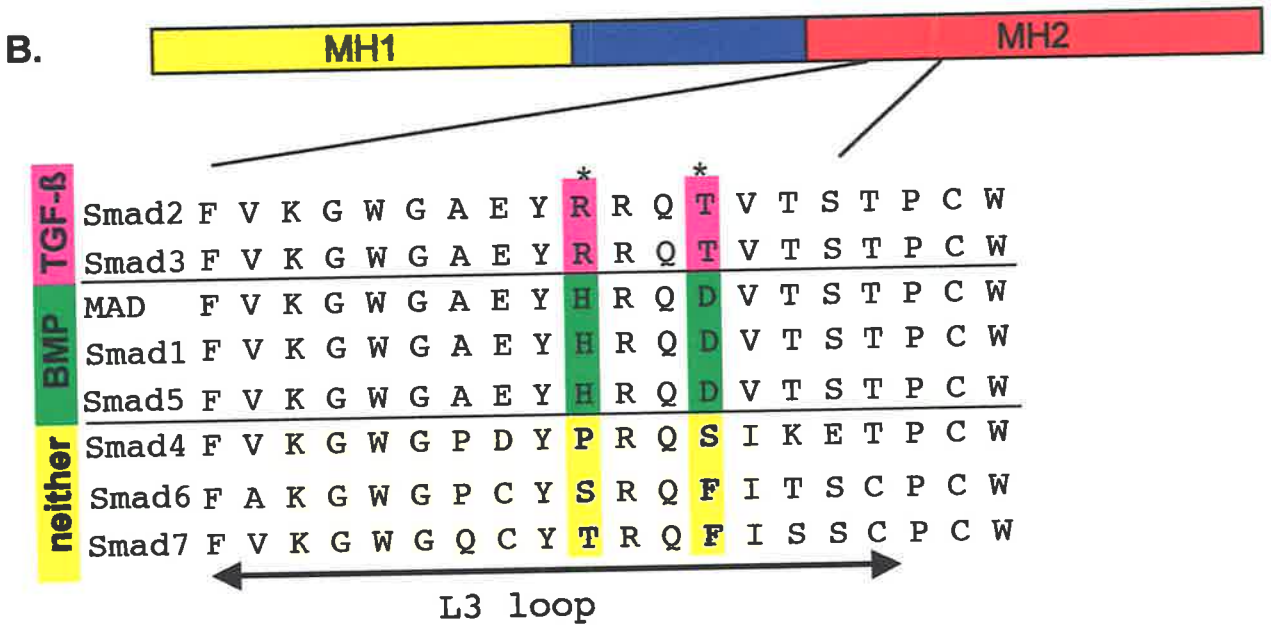
Once phosphorylated, the R-Smads dissociate from the receptor and translocate to the nucleus where they activate gene expression (Hoodless *et al.*, 1996; Macias-Silva *et al.*, 1996; Nakao *et al.*, 1997a; Nakao *et al.*, 1997b). The C-terminal domain alone

Figure 1.6. Classification of the Smads.

A: Schematic representation of the three different classes of Smads: There are three classes of Smads, the R-Smads, the co-Smad and the inhibitory Smads. Smads have two highly conserved domains, the MH1 and MH2 domains (except the inhibitory Smads which lack an MH2 domain). Red = MH1 domain; Green = MH2 domain. Note the SS/TXS motif present only in the R-Smads. The classifications of various Smads identified in *D. melanogaster, C. elegans* and vertebrates are shown. Taken from Zhang ad Derynck, 1999.

B: Alignment of the L3 sequences from various Smads: The L3 loop is within the Smad MH2 domain. Amino acids within the loop are indicated. * represents the TGF-β sub-class-specific amino acids. Note that Smad2 and 3 are TGF-β/activin-specific, and Smad1 and 5, and MAD, are DPP/BMP2/4-specific. Taken from Lo *et al.*, 1998.


A.
R-Smad
SS/TXS
co-Smad
Inhibitory Smad
Vertebrate
Smad1 Smad2
Smad3 Smad5
Smad8
Smad4
Smad6
Smad7
D. melanogaster MAD
MEDEA
DAD
C. elegans
SMA-2
SMA-3
SMA-4
DAF-3


B.
MH1
MH2
TGF-ß
BMP
neither
Smad2 F V K G W G A E Y R R Q T V T S T P C W
Smad3 F V K G W G A E Y R R Q T V T S T P C W
MAD F V K G W G A E Y H R Q D V T S T P C W
Smad1 F V K G W G A E Y H R Q D V T S T P C W
Smad5 F V K G W G A E Y H R Q D V T S T P C W
Smad4 F V K G W G P D Y P R Q S I K E T P C W
Smad6 F A K G W G P C Y S R Q F I T S C P C W
Smad7 F V K G W G Q C Y T R Q F I S S C P C W
L3 loop

can localise to the nucleus even in the absence of ligand induction (Liu *et al.*, 1996). Further, a C-terminal truncation of the R-Smads inhibits their ability to activate gene expression (Zhang *et al.*, 1996). These results suggest a regulatory role for the N-terminus whereby it masks the C-terminal domain prior to activation of the protein. After phosphorylation, unmasking allows entry into the nucleus and activation of gene expression.

1.6.4.3 Common (co-) Smad

During Smad signalling, activated R-Smads associate with the co-Smad. This Smad protein is termed Smad4 in vertebrates (eg. Hahn *et al.*, 1996; see Figure 1.6A) and MEDEA in *D. melanogaster* (Raftery *et al.*, 1995; see Figure 1.6A). It has a broad sequence similarity with other Smad family members but displays significant variance in all three molecular domains, suggesting distinct functional characteristics. In particular, the co-Smad does not contain the SS/TXS site of phosphorylation, implying that this Smad is not phosphorylated by the type I receptor. In agreement with this, the co-Smad does not co-immunoprecipitate with the ligand-receptor complex (Zhang *et al.*, 1996).

Inactivation of the co-Smad in *X. laevis* disables TGF-β, activin and BMP signalling (Lagna *et al.*, 1996) suggesting that this protein is a shared mediator of signalling. Further, cell culture studies suggest that complex formation occurs between the co-Smad and the R-Smads in response to ligand stimulation, and that this occurs after the R-Smads have been phosphorylated and dissociated from the receptor complex (Lagna *et al.*, 1996). This interaction is mediated via the MH2 domain of both the co-Smad and the R-Smads (Hata *et al.*, 1997). Co-Smads contain an L3 loop in their MH2 domain. The two residues that correspond to the subtype-specific amino acids in the R-Smads are invariant among the co-Smads, but different to those of the R-Smads (see Figure 1.6B). Mutational analysis has shown the L3 loop in the co-Smad to be essential for interaction with the R-Smads (Shi *et al.*, 1997).

Yeast 2-Hybrid experiments show that the co-Smad and R-Smads can form both homo- and hetero-interactions (Wu *et al.*, 1997). Although it has been suggested that the R-Smads function as monomers, interacting with a co-Smad trimer, there remains conflicting evidence (Shi *et al.*, 1997; Kawabata *et al.*, 1998) and the *in vivo* stoichiometry of the Smad complexes still remains to be determined.

Similar to the R-Smads, the C-terminal domain of the co-Smad acts as a transcriptional activator. Injection of a co-Smad C-terminal fragment causes constitutive, ligand

independent activation of a reporter gene (de Caestecker *et al.*, 1997). The N-terminal domain inhibits this function through direct physical interaction with the C-terminal domain and it is thought that upon heterocomplex formation with the R-Smads the inhibitory function of the N-terminal domain is released to allow gene activation (Liu *et al.*, 1996; Hata *et al.*, 1997). Thus, both types of Smad participate in DNA binding (Liu *et al.*, 1997), and although both the co-Smad and the R-Smads can activate gene transcription alone, these proteins synergise to induce strong gene expression (Lagna *et al.*, 1996; Zhang *et al.*, 1996).

1.6.4.4 DNA binding

Both the co-Smad and the R-Smads bind DNA via their N-terminal domain, this binding being necessary for transcriptional activation (Shi *et al.*, 1998). Several Smad DNA binding motifs have been identified. The first site, a twelve base pair sequence in the *vestigial* (*vg*) enhancer, is recognised by the *D. melanogaster* protein, MAD (Kim *et al.*, 1997). A GC-rich consensus sequence, GCCGnCGC, was deduced from alignment of this sequence with five further MAD binding sequences. More recently this sequence has been demonstrated to function during BMP, but not TGF-β, activation in mammals (Kusanagi *et al.*, 2000). A second DNA binding site, the Smad binding element (SBE) AGAC, was determined from the MH1 domain crystal structure of the TGF-β-activated Smad3 (Shi *et al.*, 1998; Zawel *et al.*, 1998). Binding to the major groove of DNA is mediated by a β-hairpin structure that protrudes from the surface of the MH1 domain. This region of the polypeptide is among the most highly conserved regions in the Smad proteins; all residues except the two at the turn of the hairpin are invariant among mammalian Smads (Shi *et al.*, 1998). Specific amino acids important in this binding include residues corresponding to Arg 74, Gln 76 and Lys 81 in Smad3.

Because the residues involved in DNA binding are so conserved, the specificity of Smad gene activation may be due to other sequence-specific DNA binding proteins. Some Smad binding partners identified include FAST-1 and –2, as well as AP-1, all of which interact with the TGF-β-specific, and not the BMP-specific Smads (Chen *et al.*, 1996; Labbe *et al.*, 1998; Zhang *et al.*, 1998; Zhou *et al.*, 1998). Smad1 interacts directly with Hoxc-8. Hoxc-8 normally represses the *osteopontin* gene, but after binding Smad1 it becomes dislodged from the DNA (Shi *et al.*, 1999). In addition, the *D. melanogaster* homeodomain protein Tinman (TIN) cooperates with MAD and MEDEA during activation of its own expression (Xu *et al.*, 1998). Analysis of a 550 bp enhancer from the *D. melanogaster labial* (*lab*) gene clearly highlights the

complex nature of BMP-specific Smad transcriptional regulation. This enhancer consists of a Homeodomain responsive element (HOMRE) and a DPP responsive element (DPPRE) (Marty *et al.*, 2001). These are genetically separate, but functionally linked to achieve transcriptional regulation. The DPPRE integrates two different inputs in two modules: a repressor site, flanked by MAD-binding sites on one side and CRE-binding sites on the other, and a 50 bp element that functions as a tissue-specific responsive element (Marty *et al.*, 2001).

1.6.4.5 Inhibitory Smads

The inhibitory Smads are the third class of Smads and act as regulators of TGF-β superfamily signalling. The class is comprised of Smad6, 7 (eg. Imamura *et al.*, 1997) and *D. melanogaster* Daughters Against Decapentaplegic (DAD) (Tsuneizuni *et al.*, 1997)(see Figure 1.6A). These Smad proteins have an MH2 but no MH1 domain. In addition, they lack the C-terminal SS/TXS phosphorylation motif present in the R-Smads. Phylogenetic analysis indicates that they comprise a distinct Smad sub-family (reviewed by Christian and Nakayama, 1999). It is known that the presence of these inhibitory Smads blocks the activity of the R-Smads (Imamura *et al.*, 1997; Nakao *et al.*, 1997c; Tsuneizuni *et al.*, 1997; Hata *et al.*, 1998), however the mechanism of their action is somewhat controversial. It was initially thought that the inhibitory Smads associated with receptor type I where they formed a stable complex, thus blocking R-Smad activation. However, recent studies suggest that the anti-type I complex is an artefact, and that the inhibitory Smad binds the R-Smad directly preventing it from associating with the co-Smad (Hata *et al.*, 1997). TGF-β, BMP and activin can all induce mRNA expression of the vertebrate Smad6 and 7 genes (Imamura *et al.*, 1997; Nakao *et al.*, 1997c; Takase *et al.*, 1998). In addition, DPP is necessary and sufficient for DAD expression in the developing wing (Tsuneizuni *et al.*, 1997). These facts imply that TGF-β signalling is controlled via a negative feedback loop.

1.6.5 The DPP/BMP2/4 sub-group

D. melanogaster DPP is 588 amino acids in length (Padgett *et al.*, 1987) and has the hallmarks of a TGF-β superfamily member, including a 40 amino acid signal sequence, a proregion, and a 102 amino acid C-terminal domain (Richter *et al.*, 1997). It has several N-glycosylation sites and 7 conserved cysteine residues. Comparison of *D. melanogaster* DPP with human BMP2 and 4 C-terminal regions shows 75% similarity, while there is 23-57% similarity with other TGF-β superfamily members (Massagué, 1992). *D. melanogaster* DPP and vertebrate BMP4 exhibit functional conservation, as BMP4 rescues *dpp* null mutations in *D. melanogaster* (Padgett *et al.*, 1993) and DPP can induce bone formation in vertebrates (Sampath *et al.*, 1993).

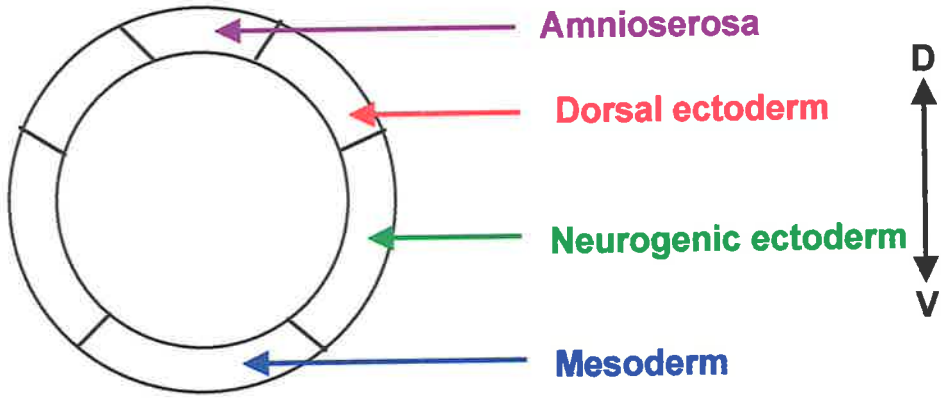


Figure 1.7. Cross-section of a simplified *D. melanogaster* blastoderm fate map showing the different regions along the dorso-ventral continuum.

Taken from Ray *et al.*, 1991.

1.7 The role of DPP in *D. melanogaster*

1.7.1 Promotion of the dorsal structures

Fate maps have allowed the identification of four distinct regions along the dorsal/ventral axis of *D. melanogaster* embryos. These regions are destined to form amnioserosa, dorsal ectoderm, ventral neuroctoderm and mesoderm (see Figure 1.7; Campos-Ortega and Vassin, 1985).

Cells specifying either amnioserosa or dorsal ectoderm are localised to the flattened dorsal side of the embryo and express *dpp* during early cellularisation. Null mutations of *dpp* result in severely ventralised phenotypes (Irish and Gelbart, 1987; St Johnston and Gelbart, 1987). Further, it has been demonstrated that embryonic cell fates along the dorsal/ventral axis are responsive to changing levels of DPP, such that high DPP activity promotes amnioserosa development while intermediate levels generate dorsal ectoderm (Ferguson and Anderson, 1992a; Wharton *et al.*, 1993). These results implicate DPP as playing a central role in the specification of embryonic dorsal fate via the establishment of an activity gradient that is highest dorsally. As *dpp* transcription is uniform throughout the dorsal region, the DPP activity gradient must be established at a post-translational level.

short gastrulation (*sog*) and *tolloid* (*tld*) are both involved in helping shape the DPP gradient. SOG is expressed in the ventro-lateral regions of the embryo (Marqués *et al.*, 1997) and diffuses dorsally where it antagonises DPP activity (Zusman *et al.*, 1988; François *et al.*, 1994). TLD is a metalloprotease, which is expressed dorsally, functioning to augment DPP activity (Arora and Nusslein-Volhard, 1992; Ferguson and Anderson, 1992a and b). TLD has been shown to cleave the SOG product in the presence of DPP (Marqués *et al.*, 1997). Consequently it appears that SOG binds DPP and limits its domain of activity ventrally. In the dorsal regions TLD cleaves SOG, releasing DPP (Marqués *et al.*, 1997).

1.7.2 Mesoderm specification

One of the first events that takes place during *D. melanogaster* development is gastrulation, or the segregation of cells into the three germ layers: ectoderm, endoderm and mesoderm. During early gastrulation, cells on the ventral side of the embryo, which express the *twist (twi)* gene, invaginate to form a layer of mesoderm that initially sits inside, and in contact with, the outer ectoderm. The mesoderm soon becomes organised into parasegmental repeats along its anterior/posterior axis, in addition to being distributed and patterned along the dorsal/ventral axis. This regionalisation appears to be coupled to the sub-division of this layer into visceral,

cardial and somatic mesoderm.

Dorsal mesodermal cells are induced by DPP, which originates in the adjacent dorsal ectoderm (Frasch, 1995; Maggert *et al.*, 1995). DPP functions to restrict the expression of the homeobox gene, *tin* (Bodmer *et al.*, 1990; Azpiazu and Frasch, 1993; Bodmar, 1993), which is initially expressed throughout the mesoderm, to the dorsal side (Staehling-Hampton *et al.*, 1994; Frasch, 1995). At stage 10 of embryogenesis, and in response to TIN, a second homeobox gene, *bagbipe* (*bap*), is activated (Azpiazu and Frasch, 1993). *bap* is not expressed throughout the dorsal mesoderm but is limited to segmental clusters of cells on the posterior side of each parasegment. Cells expressing *bap* are destined to develop into visceral mesoderm (Azipiazu and Frasch, 1993). The homeobox gene, *even-skipped (eve)* is expressed in between the *bap* cluster, on the anterior side of the parasegments, and marks the progenitors that will form pericardial cells (Frasch *et al.*, 1987). Ventral mesodermal cells not under the influence of DPP are destined to form most of the body wall somatic muscle and the fat body.

1.7.3 Midgut development

During stage 16 of embryonic development the endodermal midgut constricts at three specific points (reviewed by Bienz, 1994). DPP is secreted from cells of the surrounding visceral mesoderm, adjacent to the second of the three constrictions. Its expression is activated by the homeodomain protein Ultrabithorax (Ubx) (Capovilla *et al.*, 1994). Most prominently, DPP signalling induces the expression of the homeotic gene *lab*, in a restricted domain of the endoderm (Immergluck *et al.*, 1990; Panganiban *et al.*, 1990). The transcriptional regulation of this gene activation has already been described in Section 1.6.4.4.

1.7.4 Imaginal disc patterning: the wing as a model

Patterning of *D. melanogaster* wing discs occurs during larval development and is first apparent when the discs sub-divide into anterior and posterior compartments. DPP is expressed in a narrow band immediately anterior of the anterior/posterior border (Posakony *et al.*, 1991). Wing-specific mutations of *dpp* result in smaller or, in extreme cases, no wings. This is due to massive apoptotic cell death in mid 3rd instar wing discs (Bryant, 1988). In contrast, overexpression of *dpp* in its normal activity domain results in dramatic expansion of the wing along its anterior/posterior axis (Lecuit *et al.*, 1996). These results clearly illustrate the role of DPP in wing proliferation. Further, DPP acts cell non-autonomously, functioning at a long range throughout the wing to organise anterior/posterior pattern. (Zecca *et al.*, 1995; Nellen

et al., 1996). Analysis of the effect of patches of ectopic *dpp* expression, patches of expression of the activated DPP type I receptor, TKV, and *tkv*⁻ clones, on expression of the DPP target genes, *spalt*, *spalt-related*, and *optomotorblind*, provides evidence that this long range activity of DPP is the result of a DPP activity gradient across the wing (Zecca *et al.*, 1995; Lecuit *et al.*, 1996). TKV has been implicated in the shaping of this gradient. Highest expression of TKV is away from the endogenous source of DPP, towards the edges of the disc (Brummel *et al.*, 1994; de Celis, 1997). As a result the long-range activity of DPP is compromised by TKV which acts to sequester its ligand, limiting the spread of the gradient at the edge of the wing pouch (Lecuit and Cohen, 1998).

Later in wing development *dpp* is activated in veins where it is required for their differentiation. DPP acts here at a short-range illustrated by the high levels of TKV in the surrounding cells where they are thought to inhibit DPP diffusion (de Celis, 1997)

1.8 BMP4 in the vertebrate *X. laevis*

1.8.1 *X. laevis* BMP4 is multifunctional

Functions of BMP4 in *X. laevis* include the promotion of ventral structures. In addition, it is involved in the development of organs such as the limbs, gut, tooth, lungs and ovaries (reviewed by Hogan, 1996).

1.8.2 *X. laevis* BMP4 promotes ventral structures

At the blastula stage of development the early *X. laevis* embryo is divided into an animal and a vegetal hemisphere, between which is the marginal zone. At the start of gastrulation the marginal zone is induced to develop into mesoderm from underlying signals originating in the vegetal hemisphere (see Figure 1.8; Dale and Slack, 1987a). A second signal is released from an organising area, the Nieuwkoop centre, on the dorsal side of the vegetal hemisphere. This induces a 60^0 quadrant of the marginal zone to develop into dorsal mesoderm, also termed the Spemann organiser. During gastrulation, dorsal mesoderm will eventually form notochord while the rest of the mesoderm, the ventral mesoderm, gives rise to blood, pronephros and muscle (Dale and Slack, 1987a and b). The vegetal hemisphere develops into endoderm while the animal hemisphere forms ventral epidermis and dorsal ectoderm. This dorsal ectoderm specifies neural tissue (see Figure 1.8).

bmp4 transcripts can be identified throughout the early gastrula, bar the organiser region (Fainsod *et al.*, 1994; Schmidt *et al.*, 1995), and they have been shown to promote epidermal differentiation of ectodermal cells, and act as a neural inhibitor (Hawley *et al.*, 1995; Sasai *et al.*, 1995; Xu *et al.*, 1995). In addition, BMP4 is

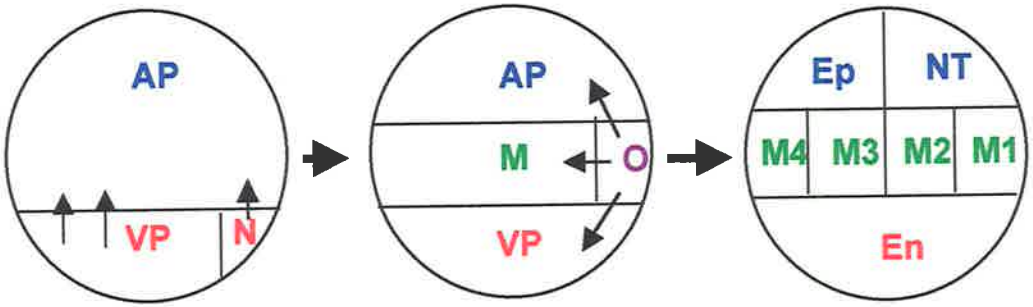


Figure 1.8. Mesoderm formation in the *X. laevis* embryo.

The vegetal pole (VP) signals to the marginal zone to induce ventral mesoderm development (M). A second signal is released from the Nieuwkoop centre (N) to induce the organiser region (O). Later, the vegetal pole forms endoderm (En) and the animal hemisphere (AP) forms the epidermis (Ep) and neural tissue (NT). The ventral mesoderm is patterned across the dorsal/ventral axis: M1 = notochord; M2 = somite; M3 = pronephros; M4 = lateral plate and ventral blood island. Taken from Dale and Wardle, 1999.

important in specifying the ventral mesoderm (Suzuki *et al.*, 1994; Hawley *et al.*, 1995). Each of the four different cell types of mesoderm (see Figure 1.8) has been shown to respond to different concentrations of BMP4 (Dale and Slack, 1987a and b; Dosch *et al.*, 1997). Ventralisation of mesoderm by BMP4 is thus concentration dependent, such that high concentrations induce a greater degree of ventralisation. Similar to DPP action, there is no evidence of a graded distribution of *bmp4* mRNA in the embryo suggesting that post-translational events are necessary for the formation of this gradient.

The dorsalising factors Noggin and Chordin have both been shown to bind BMP4 (Holley *et al.*, 1996; Zimmerman *et al.*, 1996) and these appear to be important in shaping the dorsal/ventral axis of both the mesoderm and ectoderm. Chordin is the vertebrate homolog of *D. melanogaster* SOG, sharing 47% identity.

1.8.3 Dorsal/ventral axis patterning is similar in vertebrates and *D. melanogaster*

Comparisons of the early embryonic functions of DPP and BMP4 illustrate that both genes act as morphogens to direct dorsal/ventral patterning. Although DPP is involved in dorsal patterning while BMP4 is important ventrally, it appears that the mechanism by which these genes function, namely their interactions with SOG and Chordin, respectively, has been conserved. In addition, because ventral tissues in *D. melanogaster* and dorsal tissues in vertebrates are neurogenic, DPP and BMP4 have homologous anti-neurogenic properties. These facts support an evolutionary model whereby a common ancestor had a similar mode of dorsal/ventral axis patterning, and that during the invertebrate/vertebrate divergence an inversion of this axis occurred. However, although tempting, it is important not to draw detailed analogies between vertebrate and arthropod dorsal/ventral organisation, as vertebrates have many more additional factors involved in their patterning. Furthermore, no detailed analysis of axis formation in an animal that diverged prior to the arthropod/vertebrate division has been determined.

BMP2 is an additional homolog of DPP. It is expressed at low levels during gastrulation (Clement *et al.*, 1995). Antisense injection does not cause neuralisation, so it is less likely to have an important role at this stage of *X. laevis* development.

1.9 The *D. melanogaster dpp* genetic complex

The *D. melanogaster dpp* gene is a large genetic unit of greater than 55 kb (see Figure 1.9; St. Johnston *et al.*, 1990). Five different transcripts of *dpp* have been

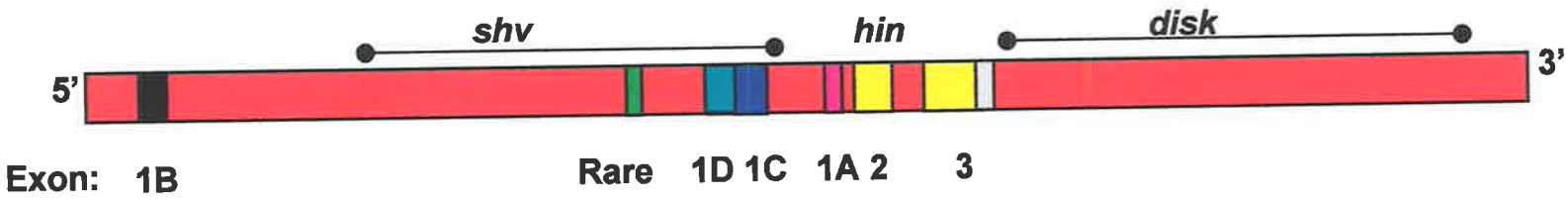


Figure 1.9. Schematic map of the *D. melanogaster dpp* genetic locus.
The three genomic regions of the *dpp* locus are marked (*shv, hin* and *disk*). Coloured boxes (bar red) represent exons. Exons 2 and 3 are common to all *dpp* transcripts and contain the open reading frame, shown in yellow. 1B, Rare, 1D, 1C and 1A represent the different exon 1 sequences present in alternative *dpp* transcripts (not drawn to scale). Adapted from Newfeld *et al.*, 1997.

identified (St. Johnston *et al.*, 1990). These range in size from 3.5 to 5.0 kb and have their own temporal and spatial pattern of expression. Each transcript consists of three exons, transcripts only differing in their 5′-most untranslated exon (reviewed by Gelbart, 1989). As the start site of the protein is located near the beginning of the common second exon, the open reading frame extending through to the common third exon, all transcripts encode an identical polypeptide (reviewed by Gelbart, 1989).

The bulk of the *dpp* gene consists of *cis*-regulatory elements controlling the timing, location and quantity of *dpp* transcription (St. Johnston *et al.*, 1990). Three genetic regions have been identified within the *dpp* complex (see Figure 1.9). The *haplo-insufficiency (hin)* region is approximately 6 kb and includes the two coding exons (St. Johnston *et al.*, 1990). *hin* null mutants are early acting embryonic lethal lesions, producing completely ventralised embryos (Gelbart *et al.*, 1985). The *disk* region is greater than 25 kb and commences just downstream of the 3′UTR of the *dpp* transcripts. Mutations in this region disrupt proper expression in imaginal discs (reviewed by Gelbart, 1989). Four out of the five 5′ untranslated exons are in the *shortvein (shv)* region. This region contains *cis*-regulatory elements driving still other spatial and temporal patterns of *dpp* expression (Segal and Gelbart, 1985; St. Johnston *et al.*, 1990) including wing vein abnormalities (St. Johnston *et al.*, 1990) and midgut malformations (Manak *et al.*, 1995).

Analysis of the genomic structure of both the human and mouse *bmp4* gene reveals a conservation of genomic organisation with their invertebrate homolog *dpp*. Both the human (van den Wijngaard *et al.*, 1996) and mouse (Kurihara *et al.*, 1993; Feng *et al.*, 1995) transcriptional units are approximately 7 kb and contain 5 exons. At least two transcripts have been detected in each organism. Each transcript contains one of either exon 1 or exon 2 followed by the third, fourth and fifth exons. The translation start is situated in exon 4 and the open reading frame extends into exon 5. The position of the splice site between the two coding exons is conserved with that of *D. melanogaster dpp*.

1.10 *A. millepora dpp*

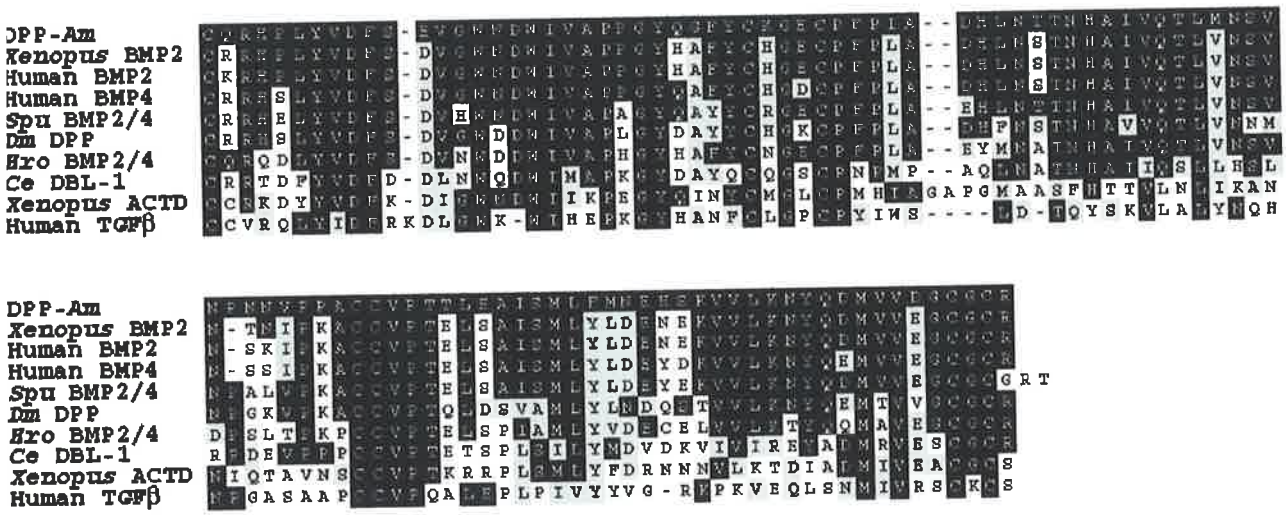
Before the commencement of this study an *A. millepora dpp* PCR product, amplified using degenerate primers, was used as a probe to screen a late embryonic stage (pre-settlement) *A. millepora* cDNA library (Brower *et al*, 1997). Four cDNAs were isolated (Hayward *et al.*, submitted). DNA sequence attained for the longest transcript

Figure 1.10. Sequence conservation of DPP-*Am*.

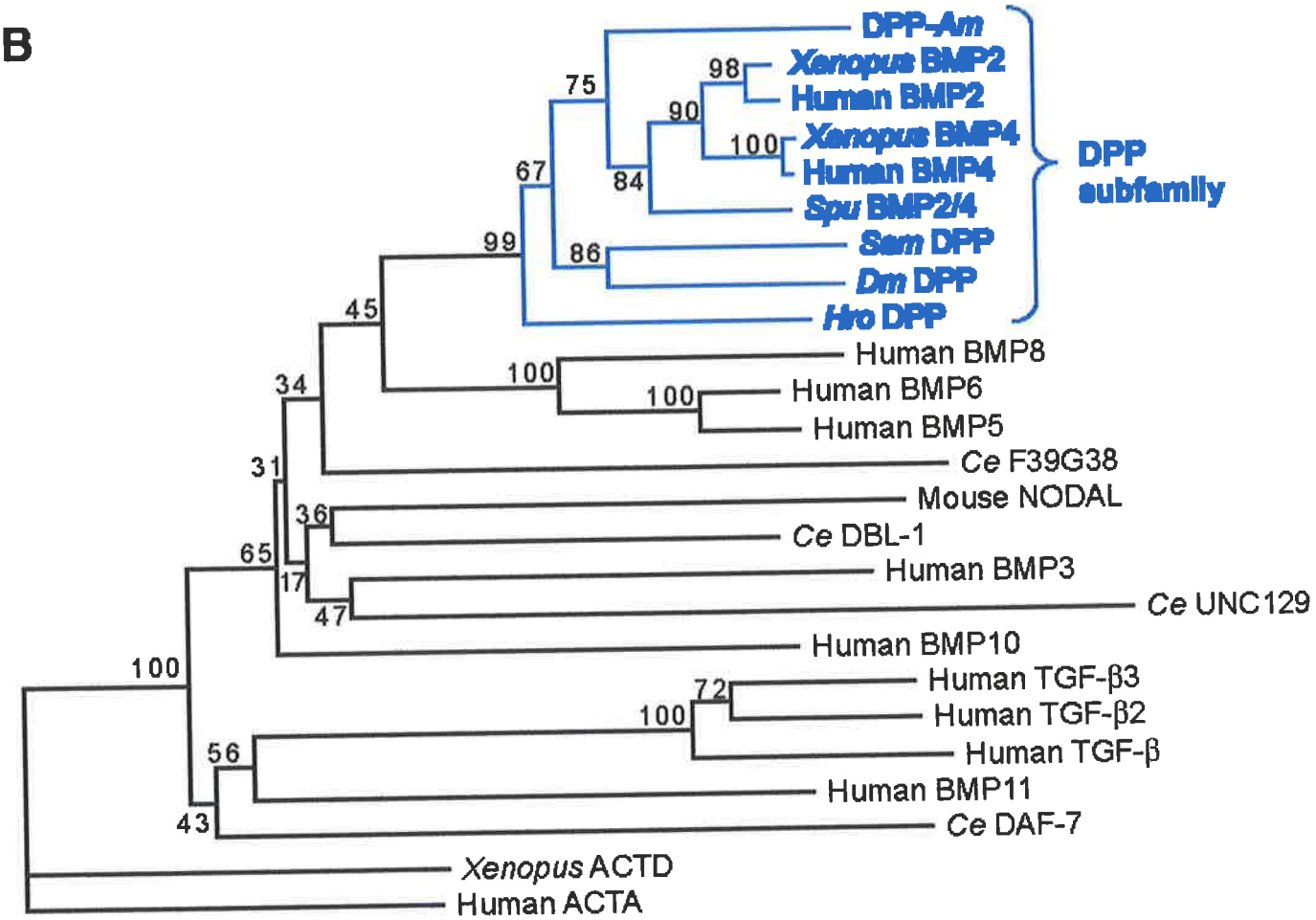
A: Alignment of the C-terminal domain of DPP-*Am* with a range of related proteins: Sequences aligned using CLUSTALW. Gaps were introduced to maximise alignment and are shown by dashes. Identical residues are boxed in black, conserved residues are boxed in grey. *Spu= Strongylocentrotu purpuratus*; *Hro= Halocynthia roretzi*; *Dm= D. melanogaster*; *Ce= C. elegans*. Taken from Hayward *et al.*, submitted.

B: Phylogenetic relationships of the C-terminal region of DPP-*Am* to those of other TGF-β family ligands: The conserved C-terminal regions of a representative range of TGF-β ligands were aligned and subjected to Neighbor-Joining distance analysis in PAUP* version 4b3 (Swofford, 2000). The human activin A sequence was defined as an outgroup. Numbers against branches indicate the percentage of 10,000 bootstrap replicates supporting the topology shown. Note that the DPP subfamily, which includes DPP-*Am*, is clearly resolved. *Spu= Strongylocentrotu purpuratus*; *Hro= Halocynthia roretzi*; *Dm= D. melanogaster*; *Ce= C. Elegans*. Taken from Hayward *et al.*, submitted.

A


DPP-Am
Xenopus BMP2
Human BMP2
Human BMP4
Spu BMP2/4
Dm DPP
Hro BMP2/4
Ce DBL-1
Xenopus ACTD
Human TGFβ

B


DPP-Am
Xenopus BMP2
Human BMP2
Xenopus BMP4
Human BMP4
Spu BMP2/4
Sam DPP
Dm DPP
Hro DPP
DPP subfamily
Human BMP8
Human BMP6
Human BMP5
Ce F39G38
Mouse NODAL
Ce DBL-1
Human BMP3
Ce UNC129
Human BMP10
Human TGF-β3
Human TGF-β2
Human TGF-β
Human BMP11
Ce DAF-7
Xenopus ACTD
Human ACTA
75
98
90
100
67
84
99
86
45
34
100
100
31
36
65
17
47
100
72
100
56
43

revealed a complete open reading frame (GenBank #AF285166). The putative protein encoded by the *A. millepora* cDNA, designated as DPP-*Am*, is most closely related to the vertebrate BMP2/4 class, matching significantly less well with *D. melanogaster* DPP (Hayward *et al.*, submitted). In the highly conserved C-terminal domain (see Section 1.6.1), sequence identity is high. For example, with *X. laevis* BMP2 it is 80% (82/102), and with *D. melanogaster* DPP it is 67% (69/102) (see Figure 1.10). Northern blot analysis with mRNA from a series of developmental stages of *A. millepora* revealed two transcripts of *dpp-Am,* and a peak in expression at the stage when the blastopore is closing (see Figure 1.11). The *A. millepora dpp* PCR product was used to screen the *A. millepora* genomic library. A single clone, *dppgen-Am*, of approximately 16 kb was isolated.

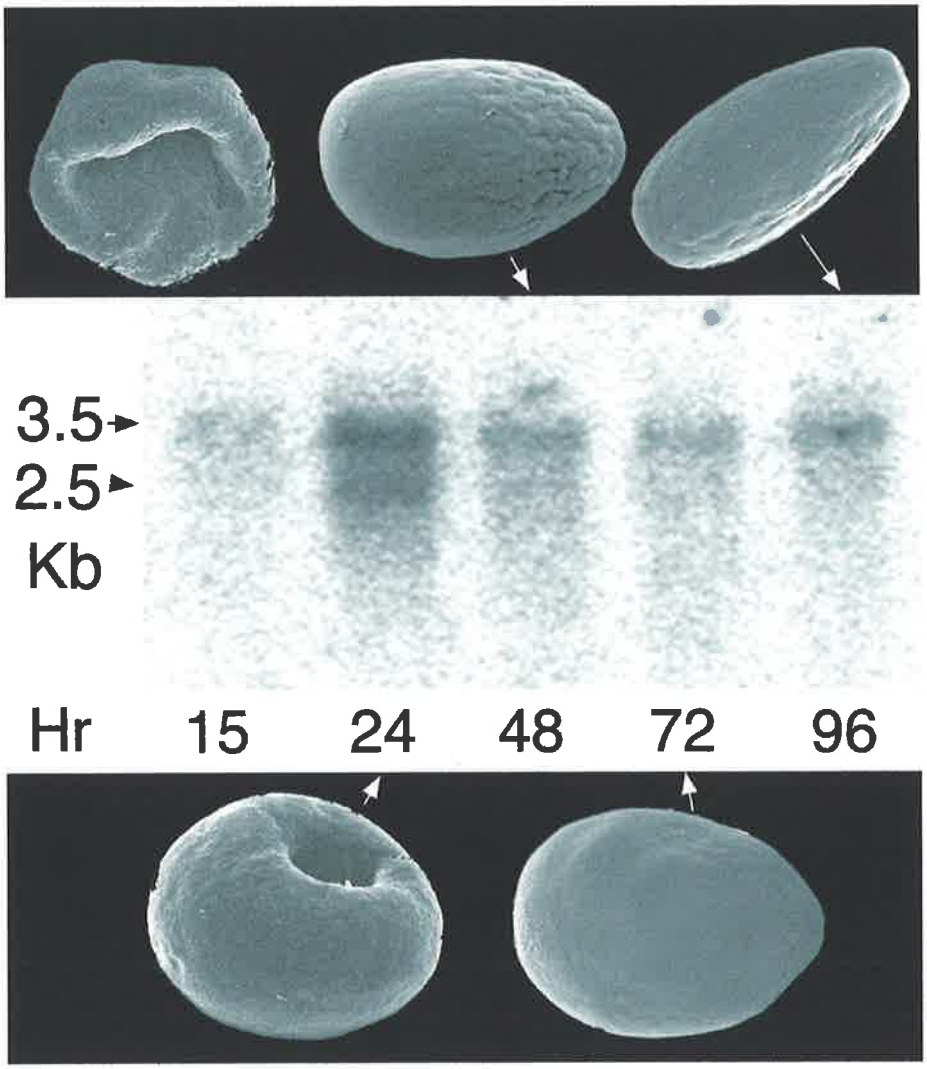
1.11 This study: Characterisation of *dpp* and other developmental genes in the cnidarian *A. millepora.*

The first section of this study undertakes a structural, functional and evolutionary characterisation of DPP-*Am* and its signal transduction cascade. Chapter three records the partial genomic sequence of the *A. millepora dpp* gene, with reflection on the evolution of this gene locus. This chapter also details the ability of *A. millepora* DPP to cause phenotypic effects in *D. melanogaster* that mimic those of the endogenous protein. Chapter four describes the isolation and evolutionary characterisation of members of the *A. millepora* DPP signalling pathway.

The purpose of the second part of this study was to expand the present knowledge regarding the molecular biology of *A. millepora* development. Specifically, chapter five details a limited *A. millepora* EST analysis. Chapter six illustrates how these analyses can and have been used to identify candidate temporal and spatial *A. millepora* marker genes, focussing on the characterisation of *A. millepora hex.*

Figure 1.11. Northern analysis of *dpp-Am*.

The morphology of the corresponding developmental stages is shown in association with a northern blot hybridised with a *dpp-Am* probe. Two transcripts, of approximately 3.5 kb and 2.5 kb, are apparent, with expression peaking at the time the blastopore is closes. Taken form Hayward *et al.*, submitted.


3.5
2.5
Kb
Hr
15
24
48
72
96

2. Materials and Methods

2.1 Materials

All chemicals and reagents were obtained from departmental stocks, being of analytical grade or the highest purity available. They were generally purchased from a range of suppliers, the major sources of the more important chemicals and reagents included Sigma, BDH Ltd, Bio-Rad, Boehringer Mannheim and Promega.

2.1.1 Enzymes

Enzymes were obtained from the following sources:

Amplitaq™: Perkin Elmer

Big Dye / Dye terminator: Perkin Elmer

Calf Intestinal Alkaline Phosphatase (CIP): Boehringer Mannheim

DNase I RNase-free: Boehringer Mannheim

eLONGase®: Life Technologies

Klenow: Amersham

Lysozyme: Sigma

Proteinase K: Boehringer Mannheim

Restriction endonucleases: New England Biolabs (NEB) and Boehringer Mannheim

Ribonuclease A: Sigma

RNase Inhibitor: Promega

T4 DNA Ligase: Boehringer Mannheim and NEB

T7/T3 RNA polymerase: Boehringer Mannheim

2.1.2 Antibiotics

Ampicillin was purchased from Sigma.

2.1.3 Radiolabelled compounds

α-^{32}P-dATP (specific activity, 3000 Ci/mmole) was purchased from Amersham.

2.1.4 Helper Phage

R408 Ex Assist Helper Phage (1.0×10^{11} pfu/ml) was purchased from Stratagene.

2.1.5 Kits

DIG RNA labelling kit: Boehringer Mannheim

HiTrap Protein G Column (1 ml): Pharmacia

Megaprime DNA radiolabelling kit: Amersham

pGEM-T® Easy Vector System 1: Promega

QIAprep Spin midiprep kit: Qiagen

QIAprep Spin miniprep kit: Qiagen

QIAquick Gel Extraction kit: Qiagen

QIAGEN Lambda Mini kit: Qiagen

QIAquick prep PCR Purification kit: Qiagen

2.1.6 Molecular weight standards

2.1.6.1 Nucleic acid

Lambda (λ) DNA, purchased from NEB, was digested with *Bst*EII and *Sal*I to produce fragments of the following sizes (in kb): 14.14, 8.45, 7.24, 5.70, 4.82, 4.32, 3.68, 3.13, 2.74, 2.32, 1.93, 1.37, 1.26, 0.70, 0.50, 0.22 and 0.12.

2.1.6.2 Protein

Pre-stained molecular weight markers were purchased from NEB and contained proteins of the following molecular weights (in kDa): 175, 83, 62, 47.5, 32.5, 25, 16.5 and 6.5.

Kaleidoscope pre-stained standards were purchased from Bio-Rad and contained proteins of the following molecular weights (in kDa): 216, 132, 78, 45.7, 32.5, 18.4 and 7.6.

2.1.7 Antibodies

2.1.7.1 Primary antibodies

The antibodies used in this thesis were provided by the following people:

anti-Even-Skipped (EVE) – polyclonal rabbit antibody kindly donated by J. Reinitz (Mount Sinai School of Medicine, New York, USA). Used at a dilution of 1:3500 in 5% foetal goat serum.

anti-Fascilin III - monoclonal mouse antibody purchased from the Hybridoma bank. Used as a three-fold dilution in PBS.

anti-Muscle Myosin - polyclonal rabbit antibody kindly donated by D. Kiehart (Duke University Medical School, Durham, NC). Used at a dilution of 1:400 in 5% Blotto.

Mab 22c10 - neural-specific mouse monoclonal antibody. Hybridoma supernatant kindly donated by N. Patel (Carnegie Institution, Baltimore, MD). Used at a dilution of 1:10 in 5% foetal goat serum.

anti-β Galactosidase - polyclonal rabbit antibody purchased from the Hybridoma bank. Used at a dilution of 1:500 in foetal goat serum.

anti-DIG-AP - obtained from Roche Diagnostics and used at a dilution of 1:2000 in PBT.

2.1.7.2 Secondary antibodies

All secondary antibodies were affinity purified, polyclonal and purchased from Jackson Immunoresearch (West Grove, PA). For immunohistochemical analysis of embryos, AP-conjugated antibodies were diluted 1:200 in 5% foetal goat serum. For western analysis the HRP-conjugated antibody was diluted 1:3000 in PBT. The antibodies included anti-mouse AP, anti-rabbit AP and anti-rat HRP.

2.1.8 Bacterial strains

All bacterial strains were present in laboratory stocks.

Escherichia coli LE392 cells were used for plating lambda phage.

E. coli BL21 cells were used for transformation and expression of protein constructs.

E. coli DH5α cells were used for propagation of plasmids and phagemids.

E. coli XL1-Blue and SOLR cells were used for excising inserts from λ ZAP II library clones.

2.1.9 *D. melanogaster* strains

w^{1118}	Laboratory stock
w ; + / *CyO* ; *Df(3R)* ro^{XB3} / *TM6B*	Laboratory stock
w; Elp/CYO TM3	Laboratory stock
w; vg::GAL4; +	Dan Kortschak
w; twi::GAL4: +	Rob Saint
w; +; UAS::*dpp-Dm*	Jules Horsfield
w; +; UAS::*prd*::GAL4/*TM3*,Sb^1 *UbxLacZ*	Julianne Camerotto
Df(3R)eR1, Ki^1/ *TM3*, Sb^1 Ser^1 (Cytological region removed: 093B0607;093D02)	Stock 3340, Bloomington Stock centre, Bloomington, Indiana

2.1.10 Plasmids

2.1.10.1 Cloning and expression vectors

pBluescriptKS+ (pKS+) - Statagene

pGEM®-T easy - Promega

pGEX-2X - obtained from Rob Saint

pMAL-c2 - NEB

pUAST (*D. melanogaster* transformation vector) - obtained from Rob Saint

2.1.10.2 Cloned DNA

Constructs were received from the following people:

pπ25.7*wc* - Rob Saint: Helper plasmid encoding Δ2-3 transposase.

pUAST-*decapentaplegic (dpp)-Am* - Dave Hayward (ANU, Canberra): 1.6 kb insert containing the *dpp-Am* open reading frame directionally cloned into the *Xho*I and *Xba*I sites of the pUAST vector.

dppgen-Am - Dave Hayward (ANU, Canberra): *dpp-Am* genomic clone.

pGEM®-1-*bagpipe (bap)* - Mount Sinai School of medicine: *bap* cloned into the *Hin*dIII and *Eco*RI sites of the pGEM®-1 vector 5' to 3', respectively.

wingless (wg) cDNA – Steve Dinardo, University of Pensylvania: *wg* inserted into pKS+.

pKS+ -*labial (lab)* 21 *Ssp*I minigene - David Miller (Indiana University, USA): *lab21* *Ssp*I minigene (2.1 kb) cloned into pKS+, such that its blunted *Ssp*I sites were inserted into the *Eco*RV site of pKS+ (Chouinard *et al.*, 1991).

ubiquitin (ubi)-CEP-52 - Dave Hayward (ANU, Canberra): A 124 bp fragment of *ubi* plus a 255 bp fragment of CEP-52 cloned into the *Eco*RI and *Not*I sites of pKS+.

2.1.11 Coral libraries

All *Acropora millepora* libraries were courtesy of Dave Hayward (ANU, Canberra).

Library	Stage	vector	Complexity
Prawn Chip	13 hours post fertilisation	λZAP II (stratagene)	3×10^8 pfu/ml
Pre-settlement	96 hours post fertilisation	λZAP II (stratagene)	1×10^9 pfu/ml
Genomic	sperm	λGEM-11 (clontech)	$5X10^5$ pfu/ml

2.1.12 Oligonucleotides

Unless otherwise stated, all oligonucleotides employed for sequence analysis or PCR amplification were obtained from Bresatec (Thebarton, SA). 5' to the left.

2.1.12.1 General oligonucleotides

T7	GTA ATA CGA CTC ACT ATA GGG C
SP6	CTA AAT CCA CTG TGA TAT C
T3	AAT TAA CCC TCA CTA AAG GG

2.1.12.2 Oligonucleotides employed to sequence *dppgen-Am*

a) Oligonucleotides kindly received from Dave Hayward

GS10	GTT AAG TGG AGC TCT GC
GS11	TCG AGC GGT CAA CAT CC
GS13	GCA CAT TTA TTA GTT GG
GS14	AGG ATC TTC TTA GTG GG
GS15	GGA GAG CCT CAA TAG CC
GS16	TGG CAG TTA TAG AGC TG
GS17	AGT TGA CTT CGC TTT CC
GS18	GTT AAG TGG AGC TCT GC

GS19	AAT TGC CTG CCA ACC GAT GG
GS21	AAG GCT ATT GAG GCT CTC CCT G
GS22	CCA AGA TGA AAA GGC ATG GA
GS23	GCA TAC TGA TGG CTT CTA GTG

b) Further primers designed for the generation of sequence data

GS25	AGG TGT TTG AGT TCA ACG TC
GS27	CAC GAG ATA TAA TCC CCA GAC
GS28	TGT CAA CCC CCA GTA TGA GG
GS29	GCT TCG ATA CGA TTT GAT GG
GS30	TAA TGC CAC CAA GCC ATG
GS31	AAC TCT GGT TCA TTG GAC TG
GS32	CAT AAG CTG TGC TCT ATC CAC
GS37	CCC TCT CTT GAC CCT AAA AAT C
GS41	CCC CGA ATG GAT GTC GAC C
GS42	GTA AAG AGC CCA CCA GGA GG
GS43	CAT TAG CCG CTT GTT GTG CC
GS44	CTC AAT CAA TGA CCG CGA GC
GS45	CGC TCC TGG CTA CTA AGC ACT GCC
GS46	GAA AGT GGC TGA GAC CC
GS47	GCG TGC GCA GCA GGA ATG CG
GS48	CCC ACG GGA CAT GGA ATC CTG CC
GS49	GAA TGA TAG CAG GTC GCC G
GS50	TGG CCG GGA AAT ATC CGA AC
GS51	GGG ACT TGC AGC CAC TGT GGC C
GS52	CAG GGA AAG ATT CGC CGC
GS53	GGC CAA GTT CGC CGG GCC ACC G
GS55	GCG CGA ACG TCC TTA AGG TG
GS56	TCT GAG GTT TCC TGC GGA TG
GS57	CGC TCA ACA CCG CAA ATA GGG G
GS58	GTT TCA GCG ATC AGC ACT ACG G
GS59	CAC CGA CAC GAA AAC ATC ACC G
GS60	CAC AAA CAC GCT GGG GGC C
GS62	GCG AGC CAC TCA GTC ATG CCG
GS64	GGC GAA TCA ATA GCG GCG CTG
GS65	GCC CAT GCC GAA TTT GAT AGC CCC
GS66	GCA ACC AAT ACA ATG CCA GCG C
GS67	ACA CGG CAA CCG CAA GAG TGA G
GS68	CTC TGG ACA TGA ATG CAC AC
GS69	AAC CCT CGA AGG ATT ACC GTC
GS70	CCA CAC ACA CCC ACC CTT CAG AC
GS71	GAC ATG GCC AAA CCA AAG CG
GS72	GCC AAA GTC AGT ACC CTA CC
GS56newrev	GCT CCC AGG TCG TTG CGG GCG GTC
GS57repeat	GAG GAG CTT GTG CTA ACC CG
GS32rev	GAG GGA CGC CCT TAC CAG TCA GTG CG

2.1.12.3 Oligonucleotides used during degenerate PCR amplification

Smad1	GCC GAA TTC CCN CAY GTN ATN TAY TG
Smad2	GCC GGA TCC CCC CAN CCY AAN ACR AA
BMP2	AAR CCN GCN ATH GCN CAY MGN GA
BMP3	ACY TCN GGN GSC ATR TAN C

2.1.12.4 Oligonucleotides used to generate protein expression constructs

a) Generation of pGEX-*smad*

Smad28REcoRI	GGA AGG ATC CAT GTC AAA TAT GTC TTC CTT G
Smad28CFBamHI	GGA AGG ATC CAT GGG AGG TGA GGT TTA TGC TG

b) Generation of pMAL-*smad*

Smad28CFEcoRI	GGA AGA ATT CAT GG G AG G TGA GGT TTA TGC TG
Smad28CRPstI	GGA ACT GCA GTT ATG CTG AGT GAG AGG TAG GCC C

2.1.12.5 Oligonucleotides used to sequence the *A. millepora* Smad cDNAs

a) Smad1/5a transcript 1

24R1	CAC TGA TAC AAA AGT CAG GAC
24R2	CTT GAC GAT GAT ACT CTG CGC CC
24R3	GCC TTC TAG TGC TCT CGA TGG TCG
24R4	GGG AGA TGG CAA AGG TGG AGG TGG
24F1	GAC TTC TAG GCT GGA AAC AGG GTG
24F2	CCC TTC AAA AGC ACA GAT GAG CCA CC
24F3	CGG TTC TAT GCT AAC AGC ACC
24F4	TGG GCC TAC CTC TCA CTC AGC

b) Smad1/5a transcript 2

28R1	GCT GAG TGA GAG GTA GGC CC
28R2	TAG TGC TCT CGA TGG TCG AG
28R3	GGT GGA GGT GGT TTT GGG C
28F1	GAC TTC TAG GCT CGA AAC AGG G
28F2	GCA ACC TTC TCC TAG CCT GAG CCC GGC

c) Smad1/5a transcript 3

26R1	TTT GCC ACC TAT CAC TGC
26R2	CTT TCT TCA ACC CTC TCT ACC C
26R3	CGA TGA TAC TCT GCG CCC CAG CC
26R4	GAC ACA TAA CAC CAT GAT GAC GGC TCC

d) Smad1/5b transcript 1

27R1	GCT CCC TAT TTG CAA GTC CCA
27R2	GAG TCA CTA AGG CAT TCA GCG
27R3	GGC ATT GGT GGA TCT TCA GCG C
27F1	CGG CCG TCT GCA GGT CTC GCA C

e) Smad1/5b transcript 2

31R1	GGA AGG ATA CAT TCC AAC ATT ATC CAC
31R2	CCA CTC AAT AGC TTG AAC AGC C
31R3	GGG TGG AAA CCA TGA CTG TGG
31R4	GTG GCG TTA CTG CAG GGG GAG
31F1	GTG CCT AAA CCA AAC CTG GGA GAC
31F2	GGC GAT GGC CCG ATC TTC AAA GCC ACC
31F3	CCA ATC AAC TAC CAG GAA CCA TCC AC
31F4	GAT TCC TTC AGG ATG CAC GC

2.1.12.6 Oligonucleotides employed to sequence the receptor cDNA

BMP 6bR1	CAC TGT CAT AAT CGC CAG CAC T
BMP 6bR2	CGG AGA TGG CGA CCT TGA TGC ACT C
BMP 6bF1	GCG AAC CAC GTT CGG CTC CGA G
BMP 1bR1	GTA CAG GAC AAC CAG ATA ACG G
BMP 6aR1	GGC CAG ATA CGC ATG TAA CGA CG
BMP 6cR1	GTC TAC TTG TCA TTC AAG GGC C
BMP 4aR1	GCA CTG GCA ACT GGT AGC AGC ACT G
BMP 5aR1	GGG ACA CAA ACA TGG CTG CTG TGA CG
BMP 1aR1	GTT TGC AAT TCT CCA TCC C
BMP 1aF1	GTA GCA CCC ACA ATT GCC CAG GGG
BMP 1aF2	GCC GAT TGA GAG CGA CAC GCA GCA GAC
BMP 4aF1	GCT ACC GAC AAC AGC ACC CCC
BMP 6cF1	GAA ACT GGA CTG TGT GAG GAG
BMP 5aF1	CTT CAT ACT GAA ATC CAG GGG
BMP 5b F1	CCA TCA AAT ACC AAC TTA CAC GGA ACC G

2.1.12.7 Oligonucleotides employed to generate Smad 3'UTR constructs

Insitu27F (smadb)	GGT GTT AAG TCA AAT GGG TTC CCC
Insitu27R (smadb)	CAA GCA GTC CCG ATA CCT TAG C
Insitu28F (smada)	GAC CAT GCC AGA AAA TGG GCC
Insitu28R (smada)	CTT CAA CCC TCT CTA CCC

2.1.12.8 Oligonucleotides employed to sequence the *A. millepora* Hex cDNA

HexAm	GGT GCT CGG GTT GAC AGA GCG GC
HexAmT7a	GGC GAC CAG CGC CAA AGC GGT ATG G
5bT3a	GTC GAT CAA CTC CAA CCA TAC CTG
5bT3b	GGT ACC ACA GAT AAT AGC CG
1aT3rev	CTC CTG TTC CAC TCG CAA TGC TGC
1aT3reva	CAT GCG ATG AGT TGA GAT CTG GTG G

2.1.12.9 Oligonucleotides employed to generate constructs of *D. melanogaster hex*

Hexinsitu1	CAG AAC CGG CGC GCC AAG TGG CG
Hexinsitu2	TCG GTG GTG GAG CTA AAG CCG CCG GTC
Hexinsitu3	GTG CCA TCG AGG CAG TGG AGC GCC
Hexinsitu4	GGC TGT TGG CTG AAC AGC GGC TGC G
HexORF1	GGA AGG TAC CGT CCC GTG GCG GTA ACA GCA GGC
HexORF2	GGA ATC TAG ATC ACT TGG GCG TCT CAT CCG CCT CGC C

2.1.12.10 Oligonucleotides used during the generation of the equalised library

Norm1	GAG ATA TTA CCT GCA GGT ACT CGT AAT ACG ACT CAC TAT AGG GC
Norm2	GAG ATA TTA CCT GCA GGT ACT CCG AAA TTA ACC CTC ACT AAA GGG
LL-SseIA	GAG ATA TTA CCT GCA GGT ACT C

2.1.13 Buffers and solutions

All solutions were prepared with distilled and deionised water and sterilised by autoclaving, except heat labile reagents, which were filter sterilised.

Agarose gel loading buffer	50% (w/v) glycerol, 50 mM EDTA, 0.1% (w/v) bromophenol blue
Blotto	5% (w/v) skim milk powder diluted in PBT
Buffer from Hell	1 M HEPES, 0.5 M EGTA, 0.1% NP-40
Denaturing solution	0.2 M NaOH, 1.5 M NaCl
Denhart's solution (5X)	2% BSA, 2% Ficoll 400, 2% PVP
10x injection buffer	50 mM KCl, 1 mM NaPO4 pH 6.8
10x MOPS running buffer	0.4 M MOPS, 0.1 M NaOAc, 0.01 M EDTA, adjust to pH 7.0 with NaOH
Neutralising solution	0.5 M Tris, 1 mM EDTA, 1.5 M NaCl
Northern pre-hybridisation solution	50% formamide, 5x Denhardts, 5x SSPE, 0.5% SDS, 20 μg/ml sonicated salmon sperm DNA
Pre-hybridisation mix	50% deionised formamide, 5x SSC, 0.5% Blotto,100 mg/ml sonicated salmon sperm DNA; 1% SDS
AP buffer	100 mM NaCl, 100 mM Tris-HCl pH 9.5, 50 mM $MgCl_2$, 0.1% Tween-20,1 mM levamisole
PBS	7.5 mM Na_2HPO_4, 2.5 mM NaH_2PO_4, 145 mM NaCl
PBT	1 x PBS, 0.1% Tween 20

Protein gel loading buffer (5x)	10% glycerol, 2% SDS, 5% β-mercaptoethanol, 0.05% bromophenol blue, 12.5%, 62.5 mM Tris-HCl pH6.8
Protein gel running buffer (5x)	1.5% (w/v) Tris-base, 7.2% (w/v) Glycine, 0.5% SDS
Re-association Buffer	O.3 M Sodium Phosphate pH 7.0, 0.4 mM EDTA, 0.04% SDS
in situ hybridisation solution (*D. melanogaster*)	50% deionised formamide, 5 x SSC, 100 µg/ml sonicated salmon sperm DNA, 50 µg/ml heparin, 0.1% Tween 20
in situ hybridisation solution (*A. millepora*)	50% formamide, 4X SSC, 1X Denhart's solution, 5 µg/ml heparin, 5% dextran sulphate, 250 µg/ml tRNA, 500 µg/ml sonicated salmon sperm DNA, 0.1% Tween 20
RIPA	150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 1 mM EDTA, 50 mM Tris pH8
20 x SSC	3 M NaCl, 0.3M sodium citrate
20 x SSPE	3 M NaCl, 0.2 M NaH_2PO_4, 0.02 M EDTA
SM buffer	0.1 M NaCl, 8 mM $MgSO_4$, 0.05 M Tris-HCl pH 7.5.
1 x SSC	150 mM NaCl, 15 mM sodium citrate
STET	50 mM Tris-HCl pH8.0, 50 mM EDTA, 8% (w/v) sucrose, 0.05% Triton X-100
1 x TAE	40 mM Tris-acetate, 20 mM sodium acetate, 1 mM EDTA, pH 8.2
1 x TBE	100 mM Tris-borate pH 7.4, 2.5 mM EDTA
1 x TE	10 mM Tris-HCl pH 7.4, 1 mM EDTA
Western transfer buffer	50 mM Tris-base, 0.3% Glycine, 0.04% (w/v) SDS, 20% methanol

2.1.14 Media

All media were prepared with distilled and deionised water and sterilised by autoclaving, except heat labile reagents, which were filter sterilised.

2.1.14.1 Bacterial media

When appropriate, the media that had been autoclaved was supplemented with Ampicilin (50 µg/ml) from sterile stock solutions.

a) Liquid media:

L-Broth: 1% (w/v) bactotyptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, pH 7.0.

SOC: 2% bactotyptone, 0.5% yeast extract, 10mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose.

b) Solid media:

L-agar plates: L-Broth with 1.5% (w/v) bactoagar, supplemented with ampicillin where appropriate.

2.1.14.2 Phage medium

Top agar: 1% (w/v) Bactotryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 10 mM MgSO4, 1% (w/v) bactoagar.

2.1.14.3 *D. melanogaster* media:

a) Fortified (F1) D. melanogaster medium: 1% (w/v) agar, 18.75% compressed yeast, 10% trecale, 10% polenta, 1.5% acid mix (47% propionic acid, 4.7% orthophosphoric acid), 2.5% tegosept (10% *para*-hydroxybenzoate in ethanol).

b) Grape juice agar plates: 0.3% agar, 25% grape juice, 0.3% sucrose, 0.03% tegosept mix.

2.2 Methods

Generally established molecular biological techniques utilised for the generation of these data have been previously published in Ausubel *et al.,* 1995 and Sambrook *et al.,* 1989.

2.2.1 Restriction endonuclease digestion

A mix containing 1 x restriction enzyme buffer, restriction enzyme and DNA was incubated at 37°C for 1-3 hours. For complete digestion, 3-5 units of enzyme were added per μg of DNA.

2.2.2 De-phosphorylation of vector DNA

2 units of CIP were added to linearised vector DNA and incubated at 37°C for 1 hour.

2.2.3 Agarose gel electrophoresis

SeaKem LE grade agarose was used for the preparation of all horizontal agarose gels. Unless otherwise specified, the agarose was melted in 1 x TAE to a final concentration of 1%. For small gels molten agarose was poured onto glass slides, whereby plastic combs were used to form well slots. For gels that were to be used for southern transfer the agarose was poured into a larger gel cast. Once set, the gels were submerged in an electrophoresis tank containing 1 x TAE. DNA samples, containing a suitable amount of agarose gel loading buffer, were loaded into the gel slots. 70 V-90 V were applied to the small gels and 150 V to the larger ones, until the dye had migrated the required distance. After staining with ethidium bromide, the gel was visualised by illumination with short wave UV, for non-preparative gels, or long wave UV, for preparative gels. The procedure above was modified during experiments analysing microarray samples, such that 400 ml of agarose, melted in 1 x TBE, was

poured into a large gel cast (SCIE PLAS). The gel was immersed into a gel tank containing 1 x TBE and, once loaded into the gel, the DNA was electrophoresed at 180 V.

2.2.4 Isolation of DNA restriction fragments from agarose gels

DNA was isolated from agarose gels by staining the gel with ethidium bromide and then excising the band of DNA under long wave UV light. The DNA was gel extracted using the QIAquick Gel Extraction kit following the suppliers' protocols.

2.2.5 Generation and transformation of recombinant plasmids

2.2.5.1 Ligation

A 10-20 µl mix containing DNA fragments to be ligated, 1 x ligation buffer and 1 unit of T4 DNA ligase was incubated at 13°C overnight.

2.2.5.2 Preparation of $CaCl_2$ competent cells

0.5 ml of an overnight culture of either *E. coli* DH5α or *E. coli* BL21 cells was used to innoculate 50 ml of L-broth. This was incubated, with shaking, at 37°C until the culture had reached an OD at A_{600} of 0.3-0.4. The cells were chilled on ice for 25 minutes, and then pelleted by centrifugation in a BECKMAN centrifuge (JA-20 rotor) at 5000 rpm for 5 minutes at 4°C. The pelleted cells were re-suspended in 25 ml of ice-cold 0.1 M $MgCl_2$, re-pelleted, re-suspended in 2.5 ml of ice-cold $CaCl_2$, and left to stand on ice for one hour. Following this, the re-suspended cells were pelleted and finally re-suspended in 12.5 ml of ice cold 0.1 M $CaCl_2$/12.5% glycerol. The competent cells were aliquoted into 1.5 ml tubes, such that each tube contained 200 µl of cell suspension, and stored at -80°C.

2.2.5.3 Transformation

a) Standard: A 200 µl aliquot of competent cells was thawed on ice for 20 minutes. DNA was then added and the cells plus DNA were incubated on ice for a further 30 minutes. Following this, the cells were heat shocked at 42°C for 90 seconds and placed on ice for 2 minutes, before 1 ml of L-broth was added to the cells and the suspension incubated at 37°C for 1 hour. Cells were then pelleted by centrifugation in a microcentrifuge for 1 minute at 14000 rpm. All bar 100 µl of the supernatent was removed and the cells were gently re-suspended in the remaining solution. The cell suspension was plated onto L-agar plates, supplemented with ampicillin, and incubated at 37°C overnight. If selection for β-galactosidase activity (blue/white colour selection) was required, 10 µl of 10% IPTG and 10 µl of 20% BCIG were added prior to plating.

b) Transformation of the normalised library: This was performed using J109 high efficiency $CaCl_2$ competent cells (Promega) according to the manufacturers' instructions.

2.2.6 Colony Cracking

For quick estimates of the size of plasmids, a rapid lysis and electrophoresis technique was employed. A bacterial colony was touched with a sterile pipette tip and dotted onto a master L-agar plate that was supplemented with ampicillin. The tip was then swirrled in a 1.5 ml tube containing 15 µl of lysis buffer (50 mM NaOH, O.5% SDS, 5 mM EDTA, 10% glycerol, 0.01% (w/v) bromophenol blue) and the tube plus tip was incubated at 65°C for 15 minutes. 10 µl of this bacterial/lysis solution mix was then dry loaded into an agarose gel, and electrophoresed at 40 V until the DNA and dye had run into the gel. The sample then underwent routine agarose gel electrophoresis.

2.2.7 Isolation of plasmid DNA

2.2.7.1 Small scale preparation-Rapid boiling lysis method

Cells from an overnight culture were harvested at 14000 rpm for 1 minute in a microcentrifuge. The supernatent was removed and the bacterial pellet was re-suspended in 200 µl of STET. 10 µl of 10 mg/ml lysozyme was added. The suspension was immediately heated at 100°C for 45 seconds, and then centrifuged in a microcentrifuge at 14000 rpm for 15 minutes. The pellet was removed with a sterile toothpick. Plasmid DNA was precipitated from the supernatant with 240 µl of iso-propanol, centrifuged at 14000 rpm for 10 minutes, and washed with 70% ethanol. The pelleted DNA was air-dried and re-suspended in 20 µl H_2O.

2.2.7.2 Preparation of ultrapure DNA

Small-scale and large-scale preparations of ultrapure DNA were prepared using the Qiagen miniprep kit and the Qiagen midiprep kit, respectively. In both instances preparations were performed according to the manufacturers' protocol.

2.2.8 Isolation of *D. melanogaster* genomic DNA

D. melanogaster genomic DNA was prepared from adult flies. 25 live flies were collected into a 1.5 ml tube and frozen on dry ice. These frozen flies were transferred to ice and homogenised in 100 µl of ice-cold 0.1 M Tris-HCl pH 9.0, 0.1 M EDTA. 100 µl of 2% SDS and 100 µg/ml proteinase K was added and the mixture was incubated at 50°C for 3 hours. Cell debris was pelleted at 14000 rpm for 10 minutes in a microcentrifuge. The supernatant was phenol/chloroform extracted twice by mixing the two liquid phases gently before centrifugation. The DNA was precipitated with 0.4 volumes of 5 M NH_4OAc and 2 volumes of 100% ethanol. The precipitate was pelleted by centrifugation in a microcentrifuge at 14000 rpm for 5 minutes at 4°C. The

DNA pellet was washed with 70% ethanol, air-dried, and re-suspended in 20 µl H_2O by placing at 37°C for 15 minutes.

2.2.9 Determination of DNA concentration

The concentration of the DNA sample was estimated by comparing it to bands of known concentration on an agarose gel, or by determining UV absorbance of the DNA solution at 260 nm, and estimating the DNA concentration (50 µg of dsDNA, and 30 µg of ssDNA, per absorbance unit).

2.2.10 PCR amplification of DNA

All PCR amplifications were performed using an MJ Research Pelteir thermal cycler.

2.2.10.1 Standard PCR

Perkin-Elmers' Amplitaq™ polymerase was employed for PCR amplification. All components for the PCR reaction were mixed in a 0.5 ml tube in accordance with the manufacturers' instructions. Specifically, 500 ng simple template or 1 µg/µl complex template (genomic DNA or library) was added to a PCR mixture containing 1 x AmpliTaq™ reaction buffer, 2.5 mM $MgSO_4$, 1 µM of each primer, 200 µM dNTPs and 2.5 units/100 µl of Amplitaq™ polymerase. After an initial 94°C 2 minute denaturation step, the PCR mix was thermally cycled through 30-40 cycles using the appropriate annealing temperature and extension time, finishing with a final extension time of 10 minutes.

2.2.10.2 Degenerate PCR amplification

This was performed as described in Section 2.2.10.1, except the it was necessary to modify the PCR program to include an additional 3 cycles of {94°C for 30 seconds; 37°C for 30 seconds; 2 minute ramp to 72°C; 72°C for 1 minute} prior to the usual cycling program.

2.2.10.3 PCR amplification of DNA for EST analysis

PCR amplification was performed in 96-well plates, as described in Section 2.2.10.1, with several modifications. Specifically, a master mix containing all components in the PCR mix, bar the DNA, was aliquoted into the wells. A bacterial colony was used as a source of DNA template. Specifically, the colony was touched with a yellow tip and dotted onto a master L-agar plate supplemented with ampicillin. The tip was then placed in the PCR reaction mix and pipetted up and down to fully transfer the cells. To lyse the bacterial cells, an initial 4 minute denaturation step, at 94°C was performed. Following this the PCR mixes were thermally cycled through 35 cycles of {94°C for 1 minute; 48°C for 1 minute; 72°C for 2.5 minutes} finishing with a 10 minute extension time at 72°C.

2.2.10.4 High fidelity PCR

DNA was PCR amplified with the eLONGase® enzyme (Life Technologies) according to the manufactures' instructions. For all PCR reactions, equal volumes of buffer A and buffer B were added. The PCR reaction was thermally cycled as described in the protocol.

2.2.11 Equalisation of the inserts of a cDNA library

2.2.11.1 DNA re-association

The cDNA inserts were re-suspended in 50 µl of re-association buffer. The DNA concentration was determined using a UV spectrophotometer. Following this, the DNA was denatured for 5 minutes at 100°C, and then allowed to re-associate at 65°C for 24 hours.

2.2.11.2 ds/ssDNA separation

0.12 M and 0.48 M phosphate buffers (pH6.8) were pre-heated to 60°C in a water bath. Meanwhile, 800 mg of hydroxyapetite (HAP) (Bio-Rad) was suspended in 10 ml of 0.12 M phosphate buffer and this suspension was gently agitated for 10 minutes. The HAP was then warmed to 60°C and loaded into a water-jacketed glass column, which was connected to a 60°C water bath. The column was washed with 5 ml of 0.12 M phosphate buffer. 750 µl of 0.12 M phosphate buffer was added to the re-associated DNA and the total volume was loaded into the column. The eluate, which contained the ssDNA fraction, was collected. 800 µl of 0.48 M phosphate buffer was then loaded into the column, allowing the collection of the dsDNA, which could be used as an experimental control.

2.2.11.3 Purifying the equalised DNA

Both fractions of DNA were transferred into separate pieces of dialysis tubing, which were clipped watertight. The tubing was immersed into 5 l of 1 x TE and placed on a magnetic stirrer at 4°C, overnight. The DNA was then precipitated overnight with 0.2 volumes of 3 M NaOAc and 2.5 volumes of 100% ethanol. The precipitate was pelleted by centrifugation in a microcentrifuge at 14000 rpm for 15 minutes. The DNA pellet was washed with 70% ethanol, air-dried, and re-suspended in 20 µl H_2O

2.2.12 Radio-labelling of DNA fragments

DNA restriction fragments were labelled with α-^{32}P-dATP using the Amersham Megaprime kit, according to the suppliers' protocol. Radio-labelled DNA was separated from unincorporated nucleotides using NICK™ Columns according to the manufacturers' instructions.

2.2.13 Southern blot transfer

DNA was electrophoresed as described in Section 2.2.3. After this, the agarose gel was soaked in denaturing solution for 10 minutes, rinsed in dH_2O, and then soaked in neutralising solution for a further 10 minutes. The gel, wells facing down, was transferred to a sheet of glad wrap. A nitrocellulose filter, cut to the same size as the gel, was saturated in neutralising solution and placed in direct contact with the gel. This was subsequently covered with three sheets of similar sized 3MM Whatmann paper, the first being saturated with neutralising solution, and then with a 5 cm thick wad of paper towels. A weighted glass plate was used to provide limited pressure to the paper towels, overnight, after which the DNA was then cross-linked to the filter by UV irradiation, using a Stratagene Stratalinker. The filter was pre-hybridised, hybridised, washed, and any bound radio-labelled probe detected, as in Sections 2.2.12 and 2.2.15.

2.2.14 Northern hybridisation

Filters were pre-hybridised, hybridised, washed, and any bound radio-labelled probe detected, as in Sections 2.2.12 and 2.2.15, with the modification that northern pre-hybridisation solution was used instead of pre-hybridisation solution.

2.2.15 Hybridisation of radio-labelled DNA probe to nylon filters

Filters were pre-hybridised at 42°C in a plastic petri dish, which contained pre-hybridisation mix, for at least 1 hour. Before hybridisation, excess pre-hybridisation mix was removed from the dish, such that just enough was left to cover the filters. Meanwhile the radio-labelled DNA was denatured at 100°C for 2 minutes. This DNA probe was then added to the filters, and allowed to hybridise for at least 16 hours at 42°C, before the probe was removed. The filters were washed twice in 2 x SSC; 0.1% SDS, for at least 15 minutes each, at RT, and then twice in 2 x SSC; 0.1% SDS, for at least 15 minutes each, at 65°C. Washes were performed with gentle agitation of the filters. Filter-bound probe was detected by either autoradiography at -80°C, with an intensifier screen, or by phosphorimager detection, using a Fujix BAS 1000 phosphorimager.

2.2.16 Phage library screening

2.2.16.1 Plating phage

50 ml of L-broth, supplemented with 10 mM $MgSO_4$ and 0.2% Maltose, was innoculated with 0.5 ml of an overnight culture of *E.coli* LE392. The culture was incubated, with shaking, at 37°C, until it had reached an OD at A_{600} of 0.6-0.8. The cells were then pelleted at 3000 rpm for 10 minutes in a BECKMAN GS-15 centrifuge, and re-suspended in 5 ml of 10 mM $MgSO_4$. The cells could now be stored

for a maximum of 5 days at 4°C. To plate, the appropriate amount of phage was added to 200 µl of cells and this suspension was incubated at 37°C for 20 minutes. Following this, molten top agar was added to the cells plus phage (3 ml if plating on 40 mm plates, 8 ml if plating on 140 mm plates), the mixture was inverted, and then poured onto L-agar plates until uniformly distributed. Once set, the plates were incubated at 37°C, overnight.

2.2.16.2 Screening the library

For the first round of screening, the phage library was plated at a density of 5 x 10,000 plaques /140 mm plate. Initially, hybond-NX filters (Amersham) were placed on the plates, for a maximum of 2 minutes, to allow phage transfer. The phage DNA was then denatured, for 3 minutes, by placing the filters, plaque side up, on 3MM Whatmann that was saturated in denaturing solution. Following this, the filters were placed on 3MM Whatmann soaked in neutralising solution, for 6 minutes. The filters were air-dried for 30 minutes, UV cross-linked using a Stratagene Stratalinker, and then washed in 2 x SSC. Filters were pre-hybridised, hybridised, washed, and any bound radio-labelled probe detected, as in Sections 2.2.12 and 2.2.15. Plaques corresponding to bound probe were selected by pulling plugs of agar from the plate with a pasteur pipette, and immersing the plugs in 500 µl of SM buffer. To elute the phage into solution, the phage suspension was shaken vigorously for 30 seconds, and then left for a minimum of 4 hours at 4°C (maximum of 6 months). Phage were screened twice again: plating at a density of 500-1000 plaques per 70 mm plate for the second round screen, and at a density of 100-200 plaques per 70 mm plate for the third round screen.

2.2.17 Isolation of phage DNA

2.2.17.1 λ ZAP insert excision

For excising inserts from λ ZAP II library clones, the R408 ExAssist Helper Phage was employed, following the protocol provided by Stratagene. The DNA from the resultant bacterial clones was isolated by usual DNA preparations for plasmids.

2.2.17.2 QIAGEN Lambda Mini Kit

For purifying ultrapure phage DNA from the λGEM-11 library, the QIAGEN Lambda Mini kit was used according to the manufacturers' instructions.

2.2.18 Automated sequencing

DNA to be employed in the sequencing reaction was prepared using the Qiagen Mini Kit according to the suppliers' protocol. The sequencing reaction was performed using the ABI Prism™ Dye Teminator Cycle Sequencing Ready Reaction Kit, or Big Dye reagents (Perkin-Elmer). This was essentially as described in the manufacturers'

protocol, with the modification of using half the described amount of reaction mix. dsDNA was used as a template and primers were de-salted. Reactions were cycled through 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes in an MJ Research Pelteir thermal cycler. For sequencing directly off a phage template the protocol was modified such that 1 µg of DNA and 1 µM of primer were added to the sequencing reaction. Extension products were precipitated with 2 µl of 3 M NaOAc pH 4.6 and 50 µl 95% ethanol, by chilling at -20°C for 10 minutes. The precipitate was pelleted at 14000 rpm for 20 minutes in a microcentrifuge, washed with 250 µl of 70% ethanol, and then dried in a 65°C heating block for 1 minute. The sequencing centre at the Department of Molecular Pathology, IMVS, Adelaide was responsible for the electrophoresis and analysis of sequencing products.

2.2.19 Expression of bacterial fusion proteins

The expression plasmid of interest was transformed into *E.coli* BL21. Subsequently, a single colony was transferred into a flask containing L-broth, supplemented with ampicillin, and grown shaking overnight at 37°C. Following this, the culture was diluted 100-fold into a flask containing L-broth, supplemented with ampicillin, and incubated with shaking at 37°C until the OD at A_{600} had reached 0.6-0.8. IPTG was added to a final concentration of 1 mM to induce protein expression. Cells were then incubated, with shaking, at 37°C for a further 2-4 hours to allow the accumulation of expressed protein. After this, the bacterial cells were pelleted at 3000 rpm for 10 minutes in a BECKMAN GS-15 centrifuge. The pellet was stored at -20°C.

2.2.20 Protein gel electrophoresis

SDS-PAGE was performed using the Bio-Rad Mini-Protean II gel electrophoresis system. 12.5% bis-acrylamide gels of 0.8 mm width were prepared according to the manufacturers' instructions. The pelleted protein samples were re-suspended in 1/20th their original culture volume of 1 x fresh protein loading buffer, and then denatured at 100°C for 5 minutes. Following this, the protein samples were loaded into the gel, and electrophoresed at 180-200 V until the bromophenol blue in the sample buffer had travelled the required distance. The proteins could now be visualised by staining the gel with Coomassie Brilliant Blue. Alternatively, they could be transferred to a nylon filter by western blotting. In the former instance, stained gels were dried on a glass slide in between to pieces of cellophane saturated with 40% methanol; 10% acetic acid.

2.2.21 Western blotting

Western blotting of proteins onto a nylon membrane was performed as described by Harlow and Lane (1988). Specifically, after separation of the proteins by SDS-PAGE, the acylamide gels were soaked in western transfer buffer for 15 minutes, to prepare them for transfer. Transfer was performed using a Bio-rad Transblot SD Semi-dri transfer cell according to the manufacturers' guidelines. In general, a current of 0.3 mA per cm^3 of gel was applied for 30 minutes. After transfer, the filter was blocked for 1 hour in 5% Blotto, and then incubated with primary antibody overnight, at 4°C. The filter was initially rinsed 3 times in PBT. This was followed by 3 x 15 minute PBT washes. Secondary antibody incubations were for 2 hours at RT, and were followed by washes the same as those for the primary antibody. The secondary antibodies were HRP-conjugated (Jackson) and detected by colorimetric detection using DAB (Sigma) staining according to the suppliers' protocol.

2.2.22 Generation and purification of antiserum

The protein of interest was separated by SDS-PAGE and visualised by staining the gel with Coomassie Brilliant Blue prepared in water (Harlow and Lane, 1988). The appropriate band of protein was excised from the gel and homogenised in an equal volume of PBS by passage through gradually narrower needles. The gel slurry was mixed with adjuvant and administered to rats (courtesy of the IMVS, Adelaide, SA). Approximately 10 μg of antigen was given to each rat at each injection. Antiserum was purified by Protein-G chromatography using a 1 ml HiTrap Protein-G column, as described in the manufacturers' instructions.

2.2.23 Fly maintenance

All flies were maintained in vials containing fly media in humidified, constant temperature rooms set at either 18°C or 25°C.

2.2.24 Collection and fixation of *D. melanogaster* embryos

Embryos, collected on grape juice agar smeared with yeast, were placed in an embryo basket and washed thoroughly with 0.7% NaCl; 0.15% Triton X-100. The basket was transferred into a dish containing 100% commercially available bleach (2% sodium hypochlorite), and washed for 1 minute 20 seconds to de-chorionate the embryos. The embryos were once again washed thoroughly with 0.7% NaCl; 0.15% Triton X-100 until all the bleach had been removed. Embryos were then transferred to a glass scintillation vial containing a two-phase mix of 5 ml heptane and 5 ml aqueous phase (4% formaldehyde; 1 x Buffer from Hell). The vial was shaken on an orbiting platform for 20 minutes, such that the interface between the liquid phases was disrupted and the embryos were bathing in an emulsion. The bottom aqueous phase was drawn off

and replaced with 5 ml of methanol, and the vial was shaken vigorously for 30 seconds to de-vitellinise the embryos. De-vitellinised embryos sink from the interface and were collected from the bottom methanol phase into a 1.5 ml tube. Embryos were rinsed 3 x times in methanol, at which point they were either processed for staining or stored at -20°C.

2.2.25 Whole mount RNA *in situ* hybridisation to *D. melanogaster* embryos

2.2.25.1 Generation of digoxygenin (DIG)-labelled RNA probes

Constructs containing cDNA inserts to be labelled were digested with an appropriate restriction enzyme. Specifically, to make antisense RNA, the enzyme had to cut the DNA insert at the 5' end of the coding sequence. A DIG RNA labelling kit was used to incorporate DIG-11-UTP into the RNA product of an *in vitro* run-off transcription reaction, according to the manufacturers' protocol. The amount of incorporated DIG-11-UTP was visualised on a dot blot. Essentially, 1 µl of each of a series of dilutions of the DIG-labelled RNA was dotted onto a nylon membrane, along with similar dilutions of an RNA probe previously quantified. The filter was UV cross-linked using a Stratagene Stratalinker, rinsed briefly in PBT, and incubated with anti-DIG-AP for 2 hours at RT. The filter was washed 3 x 15 minutes in PBT and bound antibody was detected colourimetrically as described in Section 2.2.25.3.

2.2.25.2 Preparation of embryos for hybridisation

All rinses and washes were in a 1 ml volume and carried out under gentle agitation in a 1.5 ml tube. After fixation (see Section 2.2.24), embryos were re-hydrated by passing them through the following series of washes: 70% methanol/30% PP (540 µl formaldehyde in 5 ml PBT), 50% methanol/50% PP, 30% methanol/70% PP, each for 5 minutes, and then 100% PP for 20 minutes. The hydrated embryos were further washed for 4 x 20 minutes in PBT. Following this, the embryos were firstly washed in 50% PBT/50% hybridisation solution, and secondly in 100% hybridisation solution, each for 20 minutes. The embryos were then transferred to 55°C and incubated in hybridisation solution for at least 1 hour, still. The 200 µl of DIG-labelled RNA probe was denatured at 100°C for 5 minutes and then immediately placed on ice before being added to the embryos. Hybridisation proceeded at 55°C overnight. The following day the probe was drawn off. The embryos were washed, still at 55°C for 20 minutes each with hybridisation solution, then with 50% hybridisation solution/50% PBT, and finally with PBT alone. The embryos were then brought to RT and washed, with agitation, for 4 x 20 minutes in PBT. Following this, the embryos were incubated with the anti-DIG-AP antibody for 2 hours at RT. Embryos were once

again washed for 4 x 20 minutes with PBT. The location of the anti-DIG antibody was detected colourimetrically.

2.2.25.3 Colour detection of antibody

Embryos were rinsed 4 times in AP buffer, and to the fourth rinse 3.5 µl BCIP solution (50 mg/ml in 100% DMF) and 4.5 µl NBT (100 mg/ml in 70% (v/v) DMF) were added per 1 ml of buffer. The colour was allowed to develop in the dark until the staining had proceeded far enough (as assayed on a dissecting microscope). To stop the colour reaction embryos were rinsed twice, for 5 minutes each, in PBT.

2.2.25.4 Embryo mounting, microscopy, photography and image manipulation

Mounting was in 80% glycerol; 1 x PBS. To visualise the stain, embryos were mounted onto a glass slide, under a coverslip supported by two pieces of double-sided tape. The coverslips were sealed to the glass using commercially available clear nail varnish. Embryos were then viewed using a Ziess Axiophot light microscope with Normarski optics, photographed using FUJIX *Photograb*® (Copyright 1992 FUJI PHOTO FILM CO., LTD.) version 2.0.1U photograb, and image manipulations were performed using Adobe® Photoshop® (© 1989-2001 Adobe Systems Inc.) version 6.0.1.

2.2.26 Whole mount immunostaining of *D. melanogaster* embryos

All rinses and washes were in a 1 ml volume and carried out under gentle agitation in a 1.5 ml tube. After fixation (see Section 2.2.24), embryos were firstly washed for 2 x 5 minutes, and secondly for 2 x 20 minutes, in PBT. Embryos were then blocked in 100 µl of PBT containing 5% foetal goat serum for at least 30 minutes. The blocking solution was removed and embryos were incubated with the primary antibody, with gentle agitation, at 4°C overnight. The next day embryos were warmed to RT, the antibody solution was removed, and the embryos were washed in PBT (several changes of buffer over a 1 hour time period). The embryos were then incubated with the secondary antibody for at least 2 hours at RT, with gentle agitation. Following a period of washing as for the primary antibody the location of the antibody was detected, and embryos were mounted and analysed as in Sections 2.2.25.3 and 2.2.25.4. When staining with Fascilin III there were several modifications to this procedure. Embryos were washed in PBS, still, on ice, in a volume of 1.5 ml. No blocking agent was added before the primary antibody incubations, and an additional 2 x 20 minute washes were performed following primary antibody incubations. Washes following secondary antibody addition were in PBS;0.1% NP-40, rotating at 4°C.

2.2.27 *D. melanogaster* embryo cuticle preparations

Embryos were collected and de-chorionated as in Section 2.2.24. Fixation of embryos was in 1 glycerol/4 acetic acid, for 1 hour at 60°C. Following this, embryos were mounted in Hoyers medium (30 g gum arabic, 200 g choral hydrate, 20 g glycerol in 50 ml water) under a coverslip. Vitelline membranes were popped by placing light pressure on the coverslip, the extent of pressure being monitored under the Ziess Axiophot light microscope. The coverslip was sealed with commercially available nail vanish, and embyros were cleared overnight (or for a few days) at 60°C. Embryos were viewed and pictures taken as in Section 2.2.25.4.

2.2.28 Wing preparations

Flies were collected and immersed in xylene for at least one hour (maximum a few days) before being placed under a dissecting microscope. The fly wings were then removed with forceps and covered with 40 µl of Canada balsam ductile (MERCK). A coverslip was placed over the wings and the edges of the coverslip were sealed with commercially available nail vanish. Wings were viewed and analysed as in Section 2.2.25.4.

2.2.29 Hoechst staining of wing discs

Wing discs were dissected in 1 x PBS and placed in a 1.5 ml tube, on ice. Dissected discs were then fixed for 30 minutes at RT in 4% formaldehyde; 1x buffer from hell, with gentle mixing. After fixation the discs were briefly washed 4 times in 1 ml of PBT. To the final wash, 1 µl of hoechst 33258 (10 mg/ml) was added. The tube was covered with aluminium foil and gently agitated for 1 minute. The discs were briefly washed 3 more times in PBT, before being mounted in 80% glycerol; PBS. The discs were viewed and analysed as in Section 2.2.25.4.

2.2.30 *P*-element mediated transformation of *D. melanogaster*

2.2.30.1 Micro-injection of embryos

High purity DNA for injection was prepared using the Qiagen miniprep kit, as described in Section 2.2.7.2. An injection mix was prepared to a concentration of 0.5-1 µg/µl transformation vector DNA and 0.3 µg/µl pπ25.7*wc* (Δ2-3 transposase) plasmid, in 1 x embryo injecting buffer. w^{1118} flies were allowed to lay eggs on grape juice plates at 25°C for 30 minutes prior to collection. Once collected, the embryos were de-chorionated in 100% commercially available bleach (2% sodium hypochlorite) for 1 minute and 20 seconds, and rinsed thoroughly with 0.7% NaCl; 0.15% TritonX-100. Embryos were aligned on a strip of non-toxic rubber glue such that their posterior ends would face the needle. The embryos were covered with a drop of light paraffin oil and the slide was placed on the stage of the microscope. A

micromanipulator was used to position the needle and injection was undertaken by moving the microscope stage to bring the embryos to the needle, such that a very small amount of DNA was injected into the posterior of the embryo.

2.2.30.2 Screening for transformants

After injection, the embryos were surrounded by a circle of yeast and kept moist at 25°C overnight. Once hatched, larvae were collected with thin strips of 3MM Whatman paper, and carefully placed into fly vials. Emerging adults were crossed to w^{1118} flies (virgins or males), and transformed lines were identified amongst the progeny by the Pw^{+} mini-gene eye pigmentation phenotype.

2.2.30.3 Chromosome mapping of the *P*-element integration event

Numerous independent transformants were mapped to determine the chromosome of insertion. Specifically, transformants were crossed to the doubly balanced stock *w ; +/Cyo ; Df(3R)ro*^{XB3} */ TM6B* and in the next generation male transformant flies carrying the *CyO* and *TM6B* chromosomes were selected, and crossed back to w^{1118} virgins. The progeny of this cross were scored to determine whether the *P*-element insert was segregating from the second or third chromosome. If the *P*-element was on the X chromosome then no male flies with coloured eyes would be detected amongst the progeny. Once the chromosome of insertion was determined, stable lines were generated by homozygosing the *P*-element insert, or if it was lethal, the insertion was maintained over a balancer chromosome such as *CyO* or *TM6B*.

2.2.31 RNA *in situ* hybridisation to *A. millepora* embryos

All washes were in a 2 ml volume under gentle agitation, unless stated otherwise. Fixed embryos (courtesy of Eldon Ball) were allowed to warm to RT, before being washed, for a couple of minutes each, firstly in 70% methanol/30% PBT, and secondly in 50% methanol/50% PBT. They were then transferred to a small petri dish containing PBT and dissected, as required, under a dissecting microscope. Following this, the embryos were transferred into a 2 ml tube containing RIPA and incubated at 4°C overnight. The embryos were warmed to RT and washed twice with PBS, for 5 minutes each. The PBS was then gradually replaced with 100% ethanol via a series of 5 minute washes (50% ethanol; 70% ethanol; 90% ethanol; 100% ethanol). Following this, the embryos were washed for 10 minutes in 50% ethanol: 50% xylene before being incubated for 2 hours at RT in 100% xylene. A series of 5 minute washes were performed to replace the xylene with PBT (50% ethanol/50% xylene; 100% ethanol X 2; 25% ethanol/75% PBT; 10% ethanol/90% PBT) followed by 3 PBT washes, for 5 minutes each. The embryos were gradually moved into hybridisation solution via a 10 minute wash in 50% PBT/50% hybridisation solution,

followed by a 10 minute wash in hybridisation solution alone. Fresh hybridisation solution was added and the embryos were transferred to 55°C and left, without agitation, for at least an hour, before being incubated with the denatured RNA probe (see Section 2.2.25.1) at 55°C for 48 hours, still. The embryos were then washed at 55°C in hybridisation solution for 1 X 5 minutes; 1 X 30 minutes; 1 x 12 hours and 2 x 45 minutes, before being cooled to RT and washed in 50% PBT/50% hybridisation solution for 45 minutes. The embryos were washed 3 times for 15 minutes each, in PBT, before being blocked in 250 µl 5% goat serum in PBT for 30 minutes, still. This was followed by addition of 100 µl anti-DIG-AP antibody. Incubation proceeded without agitation for 2 hours, followed by 6 PBT washes, for 30 minutes each. The location of anti-DIG stain was detected colourimetrically as described in Section 2.2.25.3.

2.2.32 Immunostaining of *A. millepora* embryos

All washes were in a 2 ml volume under gentle agitation, unless stated otherwise. Fixed embryos (courtesy of Eldon Ball) were allowed to warm to RT before being washed, for a couple of minutes each, firstly in 70% methanol/30% PBT, and secondly in 50% methanol/50% PBT. Embryos were then transferred to a small petri dish containing PBT and dissected, as required, under a dissecting microscope. Following this, the embryos were transferred into a 2 ml tube containing RIPA and incubated at 4°C overnight. The embryos were warmed to RT and washed twice with PBS, for 20 minutes each, before being blocked in 500 µl 5% goat serum in PBT for 30 minutes. This was followed by addition of 600 µl of primary antibody. Incubation proceeded at 4°C overnight and was followed by 2 quick rinses and 6 X 30 minute washes in PBT, at RT. The embryos were then blocked as above, before being incubated for 2 hours with 300 µl of secondary antibody. Embryos were washed several times over a 2 hour period, after which the location of antibody stain was detected colourimetrically as described in Section 2.2.25.3.

2.2.33 Computer analysis

2.2.33.1 General

Access to universal information regarding genes and their respective nucleotide sequences was obtained using the BLASTX search tool (Altschul *et al.*, 1997) at the National Centre of Biotechnology Information, (www.ncbi.nlm.nih.gov). For analysis of sequence data obtained during this study DNAstar software (DNASTAR, Inc.), version 5.1, was employed.

2.2.33.2 EST analysis

Removal of vector and poor quality sequence was performed using custom written software (Dan Kortschak, Adelaide University) on a Sun Blade 100 running Solaris 8. Automated quality assessment was confirmed manually.

Accepted sequences were matched against the SwissProt+SpTrEMBL database of 14/10/2001 using the BLASTX search tool at http://www.entigen.com/ .

Automated analysis of relative sequence similarities between organisms was performed using custom written software (Dan Kortschak, Adelaide University) on a Sun Blade 100 running Solaris 8.

2.2.34 Phylogenetic analysis

Maximum likelihood analyses were conducted in MolPhy version 2.3 (Adachi and Hagesawa, 1996) using the relative substitution matrix of the Dayhoff model and the local rearrangement search mode.

2.2.35 Regulatory considerations

All manipulations involving recombinant DNA were carried out in accordance with the regulations and approval of the Genetic Manipulation Advisory Committee and the University Council of the University of Adelaide.

All manipulations involving animals were carried out in accordance with the regulations and approval of the Animal Ethics Committee and the University Council of the University of Adelaide.

3. Characterisation of *A. millepora* DPP and the *A. millepora dpp* gene locus

3.1 Introduction

In bilateral animals, members of the DPP sub-family of TGF-β signalling molecules are central to the specification of the dorsal/ventral axis. The level of identity between *D. melanogaster* DPP (DPP-*Dm*) and its vertebrate homologs BMP2 and 4 is high, and conservation of biological activity has been clearly demonstrated via elegant cross-phylum experiments using DPP-*Dm* and *X. laevis* BMP4 (Padgett *et al.*, 1993; Sampath *et al.*, 1993). This has led to the hypothesis that, despite an axis inversion between arthropods and chordates (Arendt and Nubler-Jung, 1994 and 1999; DeRobertis and Sasai, 1996), these proteins may have evolved in the context of specifying the dorsal/ventral axis in bilateral animals. Because of this, the isolation of a DPP homolog in *A. millepora* (DPP-*Am*; Hayward *et al.*, submitted; see Section 1.10), an animal lacking an overt dorsal/ventral axis, was surprising. The high similarity of DPP-*Am* to both DPP-*Dm* and vertebrate BMP2/4 prompted a closer analysis of this protein.

Since analysis of genomic sequences can be a powerful tool in learning about the evolution of higher metazoan genomes, this chapter reports the sequence of the *A. millepora dpp* genomic clone, *dppgen-Am* (see Section 1.10), and analysis of the structural organisation of the *A. millepora dpp* gene locus. Furthermore, in order to gain insight into the function of DPP-*Am* at an evolutionary level, and investigate whether this coral protein might have biological activity in bilateral animals, DPP-*Am* was ectopically expressed in the fly. This chapter presents evidence for DPP-*Am* being a functional homolog of DPP-*Dm*, such that DPP-*Am* is capable of causing phenotypic effects in *D. melanogaster* that mimic those of the endogenous protein. This is illustrated in both the *D. melanogaster* embryo and in the *D. melanogaster* wing.

3.2 Sequence analysis of the *A. millepora dpp* genomic clone

DNA was prepared from the genomic clone *dppgen-Am* (see Section 2.2.17.2). Sequence analysis was initially performed with the vector primers T3 and T7 (see Section 2.1.12.1). Further sequence information was determined via a stepwise approach, in that new primers were designed from each sequencing result (see

Figure 3.1. The *A. millepora dpp* genomic locus.

A: Organisation of the *dpp* genomic clone, *dppgen-Am*: The genomic DNA is represented as a black line. Regions corresponding to exons are boxed and labelled below the line. The red shading within the boxes indicates the coding region of *dpp-Am*. The location of introns, and the position corresponding to the 3' end of DPP-*Am,* are indicated as the nucleotide distance from the 5' end of the clone. Drawn to scale.

B: The genomic organisation of *dppgen-Am* is conserved with both *D. melanogaster dpp*, and mouse/human *bmp4*: The 4th column shows the length of *A. millepora* intron II and comparable introns in both the *D. melanogaster dpp* and human/mouse *bmp4* genetic loci. The position of the splice sites in *A. millepora* intron II is conserved throughout metazoans with respect to their position in the DPP/BMP4 protein (columns 3 and 5). The 3' splice site of *A. millepora* intron I is not conserved (column 2). However the positions of this splice site in different metazoans are within 15 bps of each other (column 2). Small letters represent intron sequence. Capital letters represent exon sequence. The ATG start of the *dpp/bmp4* coding sequence is underlined and shown in red. Amino acid residues that correspond to coding nucleotides are shown in blue.

A

B

	3' splice site of intron I	5' splice site of intron II	Intron length (kb)	3' splice site of intron II
dppgen-Am	aagGAGTCTTTGCTTGATCTAAAGGATG	TTCATCATGAAGgtgagt H E G D	0.92	ggcgATCATTTGTCGG H L S A
Mouse *bmp4*	ccccagAGACACCATG	TCCATCACGAAGgtcag H H E E	0.98	cctagAACATCTGGAGA H L E N
Human *bmp4*	ccccagAGACACCATG	TCCACCACGAAGgtcagt H H E E	1.00	cctagAACATCTGGAGA H L E N
Drosophila dpp	tagTTGCAAGCGACCATG	TTTACACACAAAgtgagt F T H K	1.74	tgcagATAGTAAAATCG S K I D

Section 2.1.12.2). Completion of this analysis revealed a 15,895 bp genomic clone. Comparison of this clone with DPP-*Am* cDNA revealed that it did not contain the entire *dpp* genomic locus, because of the absence of 5′UTR sequence. Several attempts were made to isolate this 5′UTR sequence, although all unfortunately proved unsuccessful. The clone did, however, contain sequences corresponding to both the complete open reading frame of DPP and the 3′UTR. These sequences were present as two exons, separated by an intron (intron II) of 923 bp (see Figure 3.1A). The N-terminus of the open reading frame formed one exon, designated exon II. The second exon, exon III, contained the remainder of the open reading frame plus the 3′UTR. The absence of 5′UTR indicated the presence of an additional upstream exon. Sequence analysis predicts that this is separated from exon II by a distance greater than 10 kb (see Figure 3.1A).

Sequence comparisons of *A. millepora* exon II and III with comparable exons from *D. melanogaster* and vertebrate *dpp/bmp4* loci revealed a conservation of genomic organisation. This is clearly demonstrated by the invariant splice sites of *A. millepora* intron II and comparable *D. melanogaster* and human/mouse sites, with respect to their position in the DPP/BMP4 proteins (see Figure 3.1B). In contrast to this, the *A. millepora* intron I 3′ splice site is not conserved. Nevertheless, comparison of its position with the 3′ position of comparable introns in *D. melanogaster* and vertebrates showed that all these splice sites are within 15 bps of each other (see Figure 3.1B).

3.3 *A. millepora* DPP is a functional homolog of *D. melanogaster* DPP

3.3.1 Generation of transgenic flies

To ascertain whether DPP-*Am* could function in the fly it was necessary to generate transgenic flies that could conditionally ectopically express DPP-*Am*. To achieve this, the GAL4-UAS transactivation system, which has been well documented for conditional ectopic expression in the fly, was employed (Brand and Perrimon, 1993). A construct was generated containing sequence corresponding to the open reading frame of DPP-*Am* inserted into the pUAST *D. melanogaster* transformation vector, which has multiple UAS sequences upstream of the inserted cDNA (courtesy of Dave Hayward; see Section 2.1.10.2). When a GAL4 activator is expressed in *D. melanogaster* cells, under the control of an endogenous promoter of interest, it results in the transcription of the UAS cDNA in a pattern specific to this promoter (Brand and Perrimon, 1993). The pUAST-*dpp-Am* construct was microinjected into

D. melanogaster embryos, allowing its insertion into genomic DNA via *P*-element-mediated germ line transformation (see Section 2.2.30). The pUAST vector has been specifically designed for this process. It contains transposable *P*-element repeat sequences flanking the UAS cDNA, which allow insertion of the UAS cDNA into the genomic DNA in the presence of a transposase (see Section 2.1.10.2). The presence of a white ($P[w^+]$) minigene within the transposable element, which produces eye pigment, allows easy detection of transformants.

To avoid any phenotypic effect that may be caused by the insertion site of the pUAST construct, two independent *D. melanogaster* lines carrying the *dpp-Am* transgene were analysed. The first line, designated T2, was mapped to the third chromosome and was homozygous viable. The second line, referred to as T3, was mapped to the second chromosome and was homozygous lethal.

3.3.2 Phenotype of ectopic DPP expression in the mesoderm

As described in Section 1.7.1, *dpp-Dm* is expressed in both the amnioserosa and dorsal ectoderm of the blastodermal embryo, and is essential for establishment of the dorsal/ventral axis, acting as a dorsal-specific factor (for example, Ferguson and Anderson, 1992a; Irish and Gelbart, 1987). In addition, it signals from the dorsal ectoderm to the underlying cells to specify dorsal mesoderm (Frasch, 1995; Maggert *et al.*, 1995). Ectopic expression of a UAS::*dpp-Dm* cDNA construct in mesodermal tissues using mesoderm-specific *twist (twi)*::GAL4 results in their dorsalisation (Staehling-Hampton *et al.,* 1994). To examine whether the UAS::*dpp-Am* cDNA had a dorsalising phenotype in this assay, the UAS::*dpp-Am* lines were crossed to the *twi*::GAL4 line. A UAS::*dpp-Dm* cDNA line, UAS::*dpp-Dm* (Horsfield *et al.*, 1998), was used as a positive control for the effects of induced *dpp* expression. The resulting phenotype in the *twi*::GAL4-UAS::*dpp-Am* embryos was dramatic and indistinguishable from those of the *twi*::GAL4-UAS::*dpp-Dm* embryos (see below), both cases resulting in 100% larval lethality.

To examine the basis of this phenotype, *bagpipe* (*bap*) expression was analysed as a marker of DPP function. Ordinarily, DPP induces *tinman (tin)* transcription in the dorsal mesoderm of stage 10 embryos (Azpiazu and Frasch, 1993). TIN subsequently activates *bap* in a segmentally repeated pattern in the dorsal mesoderm, which can be clearly visualised in stage 10 embryos stained for *bap* mRNA (Azpiazu and Frasch, 1993; Figure 3.2A). pGEM-1-*bap* (see Section 2.1.10.2) was linearised via restriction digest using the *Hin*dIII enzyme, and subsequently used as a template to generate a

DIG-labelled RNA *bap* probe for *in situ* experiments (see Section 2.2.25). In embryos ectopically expressing UAS::*dpp-Am* in the mesoderm, *bap* expression extended ventrally relative to its normal dorsally restricted domain (Figure 3.2). These results were consistent with those published by Staehling-Hampton *et al.* in 1994 for DPP-*Dm*, and demonstrate the ability of exogenous DPP-*Am* to have dorsalising activity in the fly embryo.

TIN also activates *even-skipped* (*eve*) expression in the dorsal mesoderm where, in stage 10 embryos, it is restricted to a subset of cells on the anterior side of each parasegment. These cells represent the pericardial cells (Frasch *et al.*, 1987). Eve-expressing cells therefore provided an additional marker that could be employed to test the dorsalising properties of DPP-*Am*. An Eve polyclonal antibody (see Section 2.1.7) was used to detect Eve protein localisation in embryos ectopically expressing UAS::*dpp-Am* (see Section 2.2.26). Figure 3.3 clearly highlights the identical effects of both ectopic DPP-*Am*, and DPP-*Dm*, expression on these *eve*-positive pericardial cells. The regular, segmentally repeated pattern of cells was replaced by a highly disorganised array whereby the dorsal migration of these cells has been affected. At later stages of dorsal closure this migrational effect is more evident, illustrated by the inability of these cells to reach their final destination (data not shown).

The mis-expression of *bap* and *eve* clearly highlights the disruption of ventral mesoderm patterning caused by ectopically expressing UAS::*dpp-Am* in the mesoderm. To analyse whether this ectopic DPP also had an effect on the patterning of the ventral-specific neurogenic ectoderm, whole mount embryos were immunohistochemically analysed using the neural marker mAb 22C10 (see Sections 2.1.7 and 2.2.26). The results illustrated a disruption of the neural pattern (Figure 3.4). However, a noticeably more obvious effect was the significant disruption to gut morphogenesis, as evidenced by the protrusion of the midgut out of its normal internal location. Visceral musculature normally constrains the position of constrictions, and subsequent folding of the midgut (reviewed by Bienz, 1994). Thus, to examine the basis of this midgut protrusion phenotype, embryos ectopically expressing UAS::*dpp-Am* were stained with anti-fasciclin III and anti-muscle myosin (see Sections 2.1.7 and 2.2.26) to visualise visceral mesoderm/musculature. Both stains reveal the absence of visceral mesoderm/musculature (Figures 3.5 and 3.6), thus predicting the absence of midgut constrictions consistent with the gut protrusion phenotype. Analysis of *wingless (wg)* expression further confirms the above mentioned visceral mesoderm/musculature defects. In stage 13 embryos *wg* transcripts

Figure 3.2. Ectopic expression of *dpp* in the *D. melanogaster* mesoderm leads to the mis-expression of *bap*.

Distribution of *bap* mRNA in stage 10 embryos: In wild-type embryos, *bap* expression is confined to the dorsal mesoderm (A). *bap* expression extends ventrally in embryos ectopically expressing either *twi::*GAL4-UAS::*dpp-Dm* (B) or *twi::*GAL4-UAS::*dpp-Am* (C). Lateral view, anterior left.

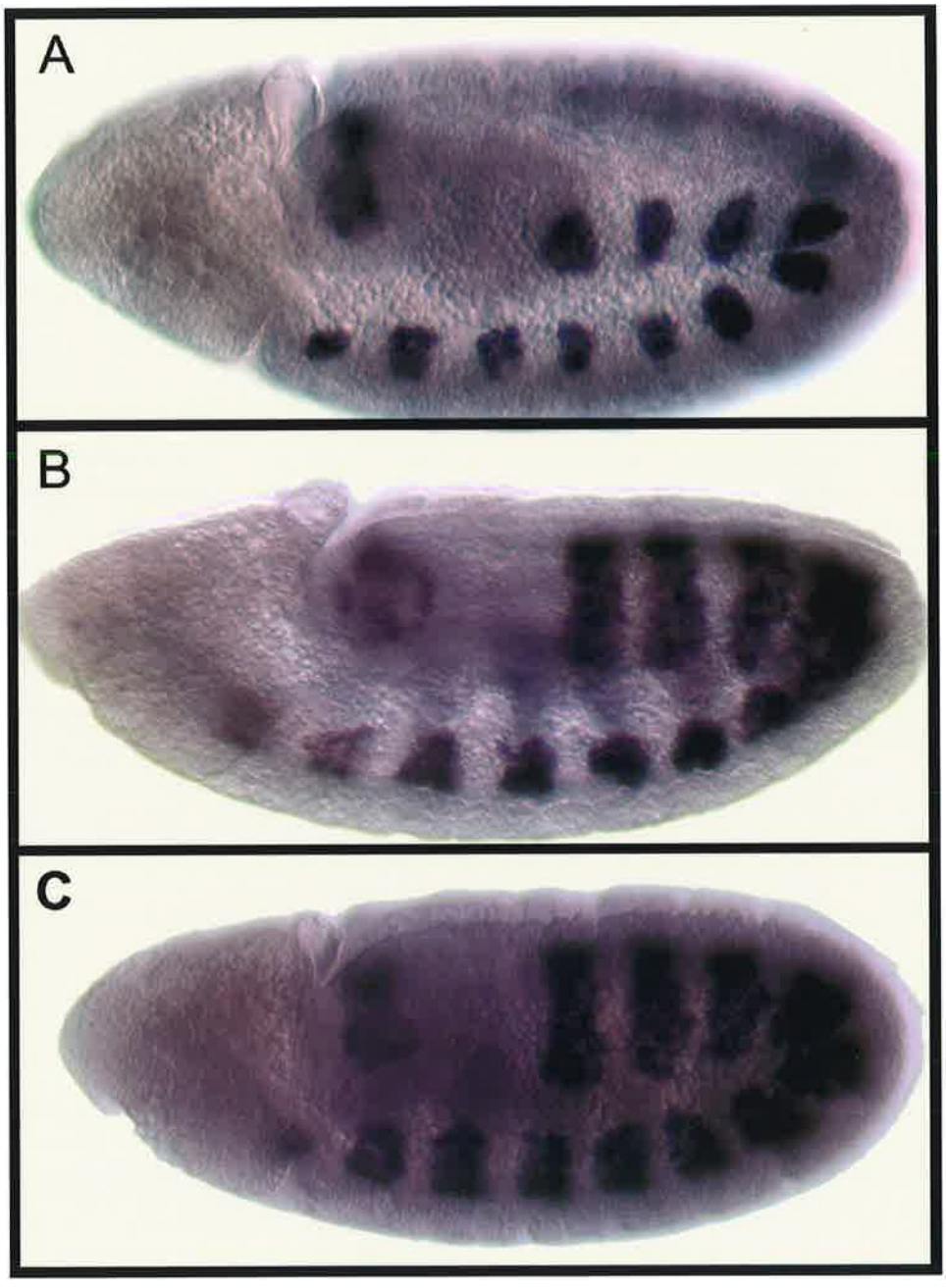

A
B
C

Figure 3.3. Ectopic expression of *dpp* in the *D. melanogaster* mesoderm leads to the mis-placement of *eve*-positive cells.

Distribution of Eve in stage 13 embryos: In wild-type embryos the Eve protein is expressed as a segmentally repeated pattern in the pericardial cells (A). *eve*-expressing cells form a disorganised array in embryos ectopically expressing *twi::*GAL4-UAS*::dpp-Dm* (B) or *twi::*GAL4-UAS*::dpp-Am* (C). Dorsal view, anterior left. A′, B′ and C′ are higher magnifications of A, B and C, respectively.

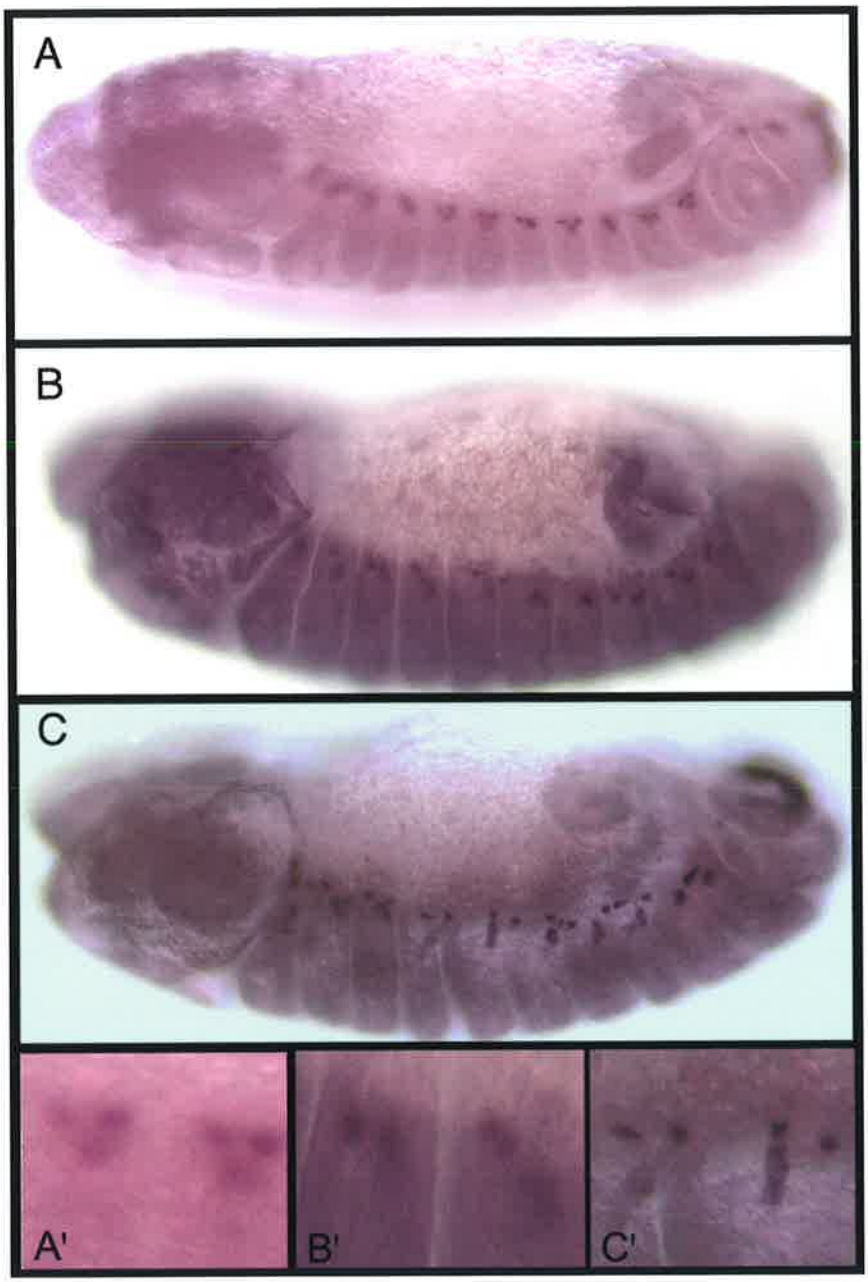
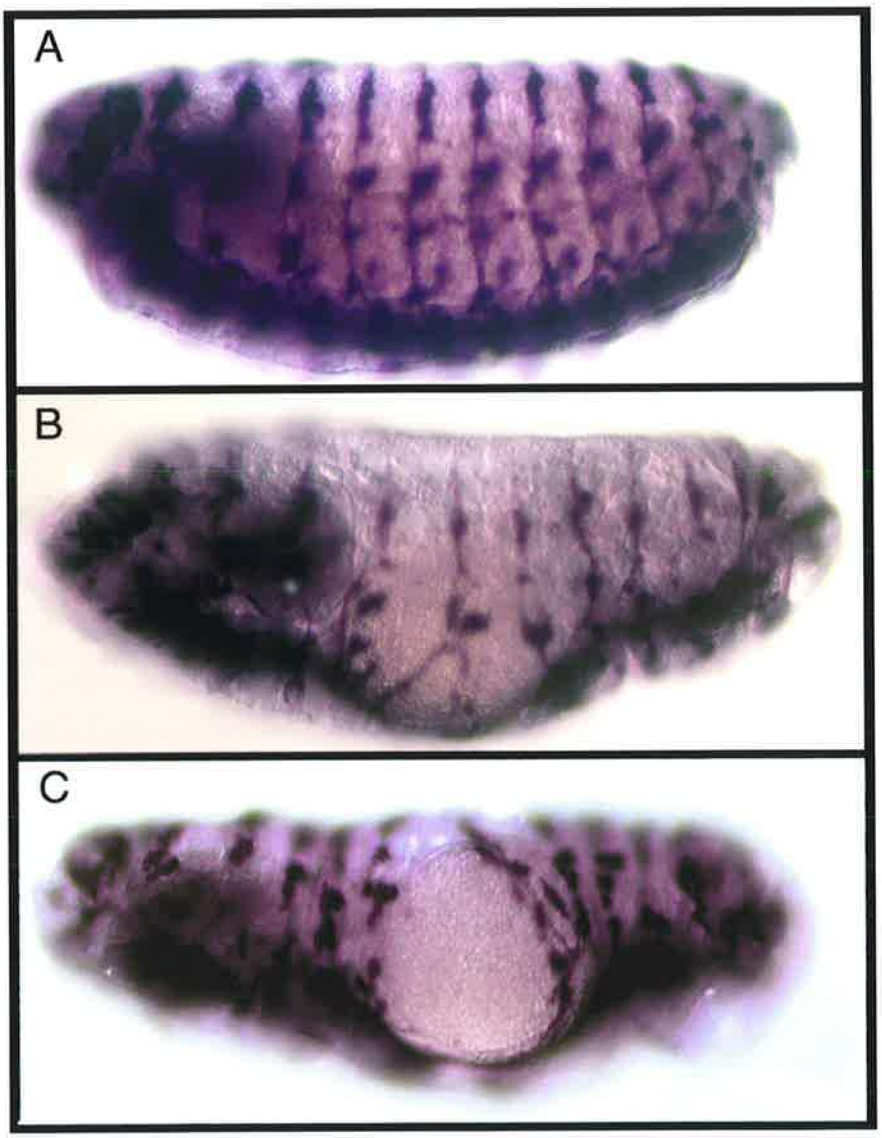

A
B
C
A'
B'
C'

Figure 3.4. Ectopic expression of *dpp* throughout the *D. melanogaster* mesoderm leads to the disruption of gut morphogenesis.

Stage 16 embryos stained with mAb 22c10: A wild-type embryo is shown in A. The midgut protrudes from its normal location in embryos that are ectopically expressing *twi::*GAL4-UAS*::dpp-Dm* (B) or *twi::*GAL4-UAS*::dpp-Am* (C). Lateral view, anterior left.


A
B
C

are ordinarily restricted to the visceral mesoderm. From here Wg signals to the underlying endoderm, and is required for the formation of the second midgut constriction (Immergluck *et al.*, 1990). Wg cDNA (see Section 2.1.10.2) was used to generate a *wg* DIG-labelled RNA probe (courtesy of Tetyana Shandala). This probe was hybridised (see Section 2.2.25) to whole mount embryos etopically expressing UAS::*dpp-Am*. Figure 3.7 reveals that *wg* expression in stage 13 embryos is lacking, consistent with the loss of visceral mesoderm.

At a first glance this loss of visceral mesoderm/musculature appears to contradict the observation by Frasch (1995) that BAP ectopic expression results in ectopic visceral mesoderm development. In his experiments, however, ectopic BAP expression was a result of ectopic *dpp* expression driven by a *Krüppel* (*Kr*)::GAL4 driver which had four ventral repression elements from the *zerknült* (*zen*) gene placed upstream. *Kr* is ordinarily expressed in the centre of the embryo, with respect to the anterior/posterior axis, and the *zen* elements resulted in the silencing of expression from the *Kr* enhancer in the ventral-most region of blastodermal embryos. Because it is this region that is destined to form mesoderm (see Figure 1.7), ectopic *dpp* was restricted solely to the ectoderm, with no mesoderm expression. In contrast to this, in the experiments described here expression of ectopic *dpp* expression was restricted to mesodermal tissues, driven by the *twi* enhancer. These results illustrate that the different drivers, although both capable of inducing ectopic *bap* expression, have quite different biological outcomes, perhaps because of differences in the level, spatial location and/or persistence of *dpp* expression.

To confirm the defects in the ventral-specific neuroectoderm illustrated by the mAb 22C10 neural marker, and to further assess the late embryonic effect of ectopic expression of DPP-*Am* compared with DPP-*Dm*, cuticles from the *twi*::GAL4-UAS::*dpp* embryos were analysed (see Section 2.2.27) to enable the visualisation of the ventral epidermis. Ordinarily, there is a distinct pattern of segmentally repeated denticle belts on the ventral epidermis of *D. melanogaster* embryos. On the dorsal side of the embryo a characteristic pattern of dorsal hairs can be seen. Both the ventral denticle belts and the dorsal hairs are initially apparent at stage 16 of *D. melanogaster* embryogenesis. Although no dorsal hairs were detected on the ventral surface of *twi*::GAL4-UAS::*dpp-Am* stage 16 embryos, there was a marked reduction in both number and size of the ventral-specific denticle belts (Figure 3.8). This was most prominent at the posterior of the embryos, where, in extreme cases no belts could be seen. This result demonstrates that, like DPP-*Dm*,

Figure 3.5. Visceral mesoderm/musculature is absent in *D. melanogaster* embryos ectopically expressing *dpp* in the mesoderm.

Stage 13 embryos stained for Fasciclin III protein: Ordinarily visceral mesoderm/musculature surrounds the prospective gut (A; arrow). There is an absence of this mesoderm/muscle in embryos that are ectopically expressing *twi::*GAL4-UAS*::dpp-Dm* (B; arrow) or *twi::*GAL4-UAS*::dpp-Am* (C; arrow). Lateral view, anterior left.

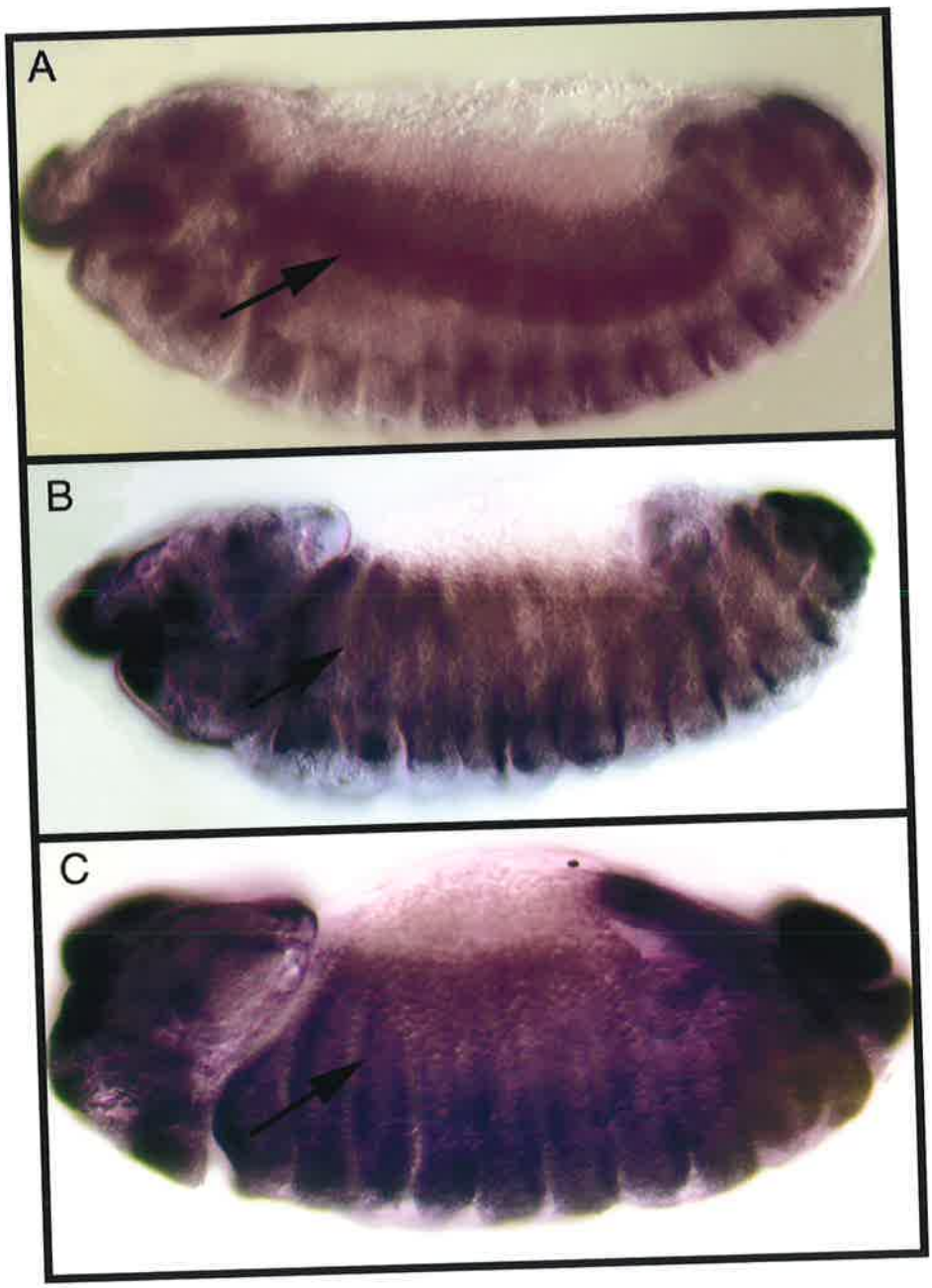

A
B
C

Figure 3.6. Visceral musculature is absent in *D. melanogaster* embryos expressing ectopic *dpp* in the mesoderm.

Stage 13 embryos stained for muscle myosin: Ordinarily visceral musculature surrounds the prospective gut (A; arrow). There is an absence of this muscle in embryos that are ectopically expressing *twi::*GAL4-UAS*::dpp-Dm* (B; arrow) or *twi::*GAL4-UAS*::dpp-Am* (C; arrow). Dorsal view, anterior left.

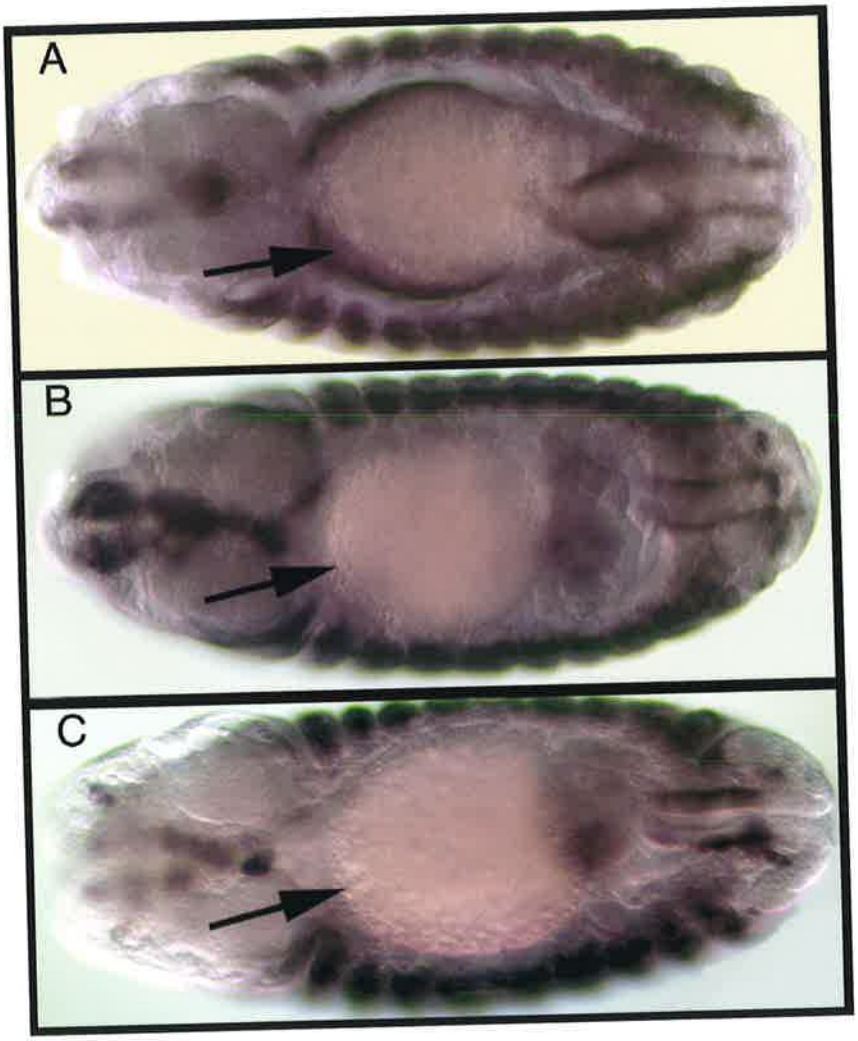

A
B
C

Figure 3.7. *wg* transcripts are absent in *D. melanogaster* embryos ectopically expressing *dpp* in the mesoderm.

Stage 13 embryos stained for *wg* mRNA: Ordinarily *wg* is expressed in the visceral mesoderm of stage 13 embryos at the position where the second midgut constriction will develop (A; arrow). No *wg* expression is seen in this position in embryos that are ectopically expressing *twi::*GAL4-UAS*::dpp-Dm* (B; arrow) or *twi::*GAL4-UAS*::dpp-Am* (C; arrow). Lateral view, anterior left.

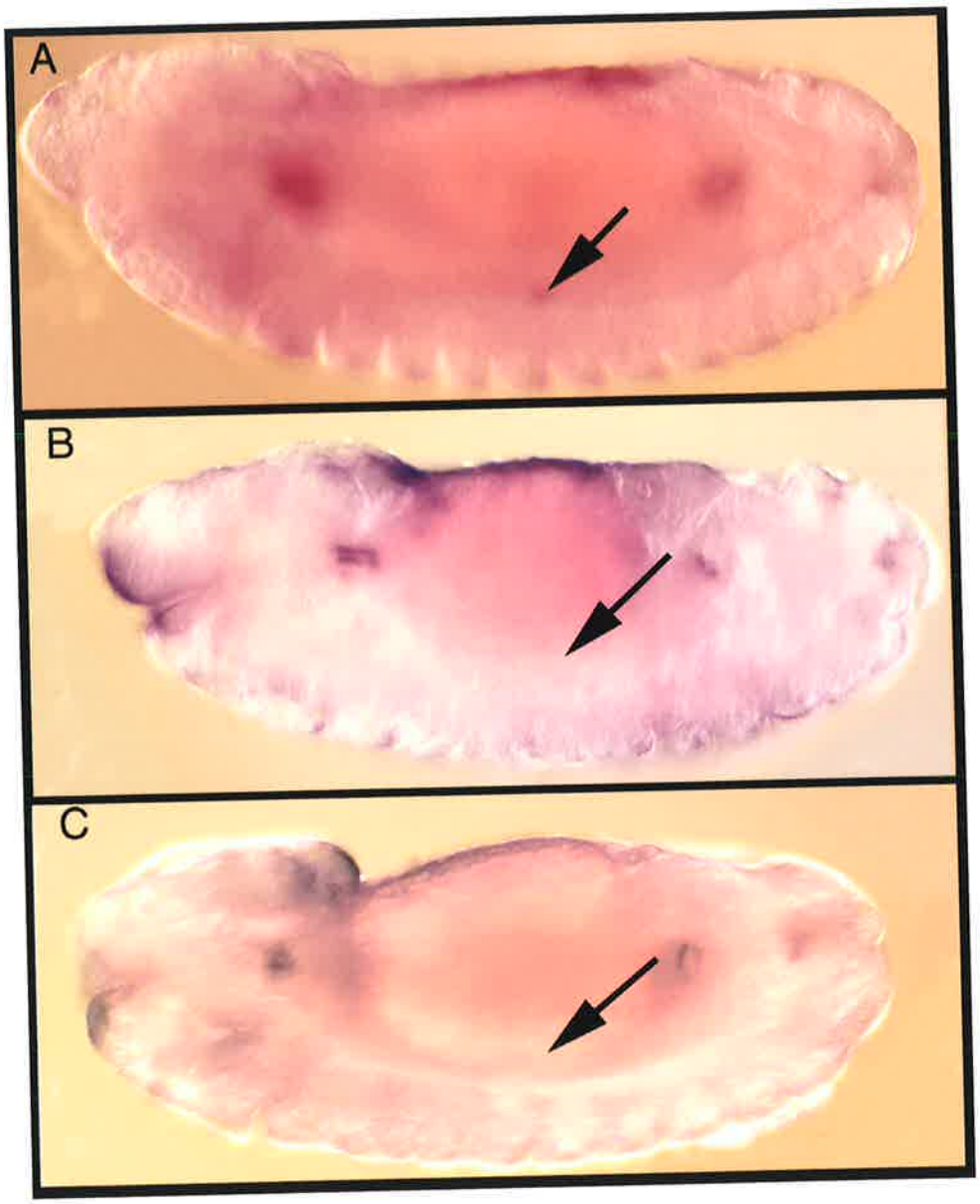
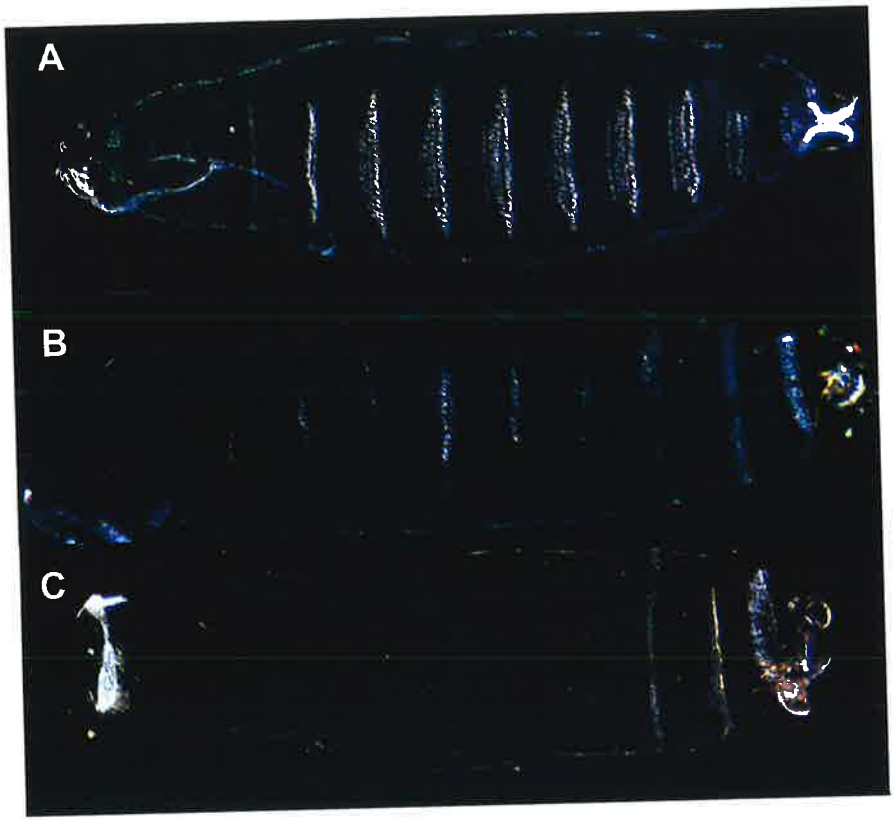

A
B
C

Figure 3.8. Ectopic expression of *dpp* in the *D. melanogaster* mesoderm affects the patterning of the ventral epidermis.

Cuticle preparations of stage 16 embryos: In wild-type embryos there is a characteristic pattern of denticle belts visible on the surface of the ventral epidermis (A). In embryos ectopically expressing either *twi::*GAL4-UAS*::dpp-Dm dpp* (B) or *twi::*GAL4-UAS*::dpp-Am* (C) these ventral denticle belts are less prominent, and sometimes absent. Ventral view, anterior left, dark field.


A
B
C

DPP-*Am* has the ability to signal across germ layers, from the mesoderm to the ventral ectoderm, where it can disrupt ventral-specific development.

3.3.3 Phenotype of ectopic DPP expression in the fly wing

Section 3.2.2 clearly illustrates that mesoderm-specific ectopic expression of both *dpp-Am* and *dpp-Dm* in *D. melanogaster* embryos results in similar phenotypes. To further these results, it was important to analyse whether ectopic expression of UAS::*dpp-Am* could mimic that of UAS::*dpp-Dm* in an alternative functional assay. To this end, *dpp-Am* was ectopically expressed in the wing. Adult *D. melanogaster* wings are formed from the wing imaginal discs, and patterning of these discs occurs during larval development. DPP expression is restricted to a narrow band immediately anterior of the anterior/posterior border of the disc (Posakony *et al.*, 1991) where it has a role in both the proliferation and the anterior/posterior patterning (Byrant, 1988; Lecuit *et al.*, 1996). Ectopic expression of UAS::*dpp-Dm* in the wing results in overgrowth which, in extreme cases, can be seen as blistering of the wing (Haerry *et al.*, 1998). To examine the possibility that DPP-*Am* might have biological activity in this aspect of DPP function, this experiment was reproduced using the two UAS::*dpp-Am* lines crossed to a wing-specific *vestigial (vg)*::GAL4 line (courtesy of Dan Kortschak). The UAS::*dpp-Dm* line was used as a positive control. Wings were prepared as in Section 2.2.28. Wings from the UAS::*dpp-Am*-T2 line had a mild phenotype. Nevertheless, there was a marked increase in the size of their posterior compartment (see Figure 3.9A). In the UAS::*dpp-Am*-T3 line, and in the UAS::*dpp-Dm* line, a more severe phenotype was produced, illustrated by a blistering of the wing (Figure 3.9A). To establish the cause of this wing blistering, wing discs from larvae ectopically expressing UAS::*dpp-Am*-T3 were analysed (see Section 2.2.29). Hoechst staining of these wing discs suggests that the blistering effect was a result of wing overgrowth (Figure 3.9B). *A. millepora* is a primitive species, which has no appendages. Despite this, DPP-*Am* is able to mimic DPP-*Dm* function in the *D. melanogaster* wing appendage.

3.4 Discussion

3.4.1 The *A. millepora dpp* genetic locus

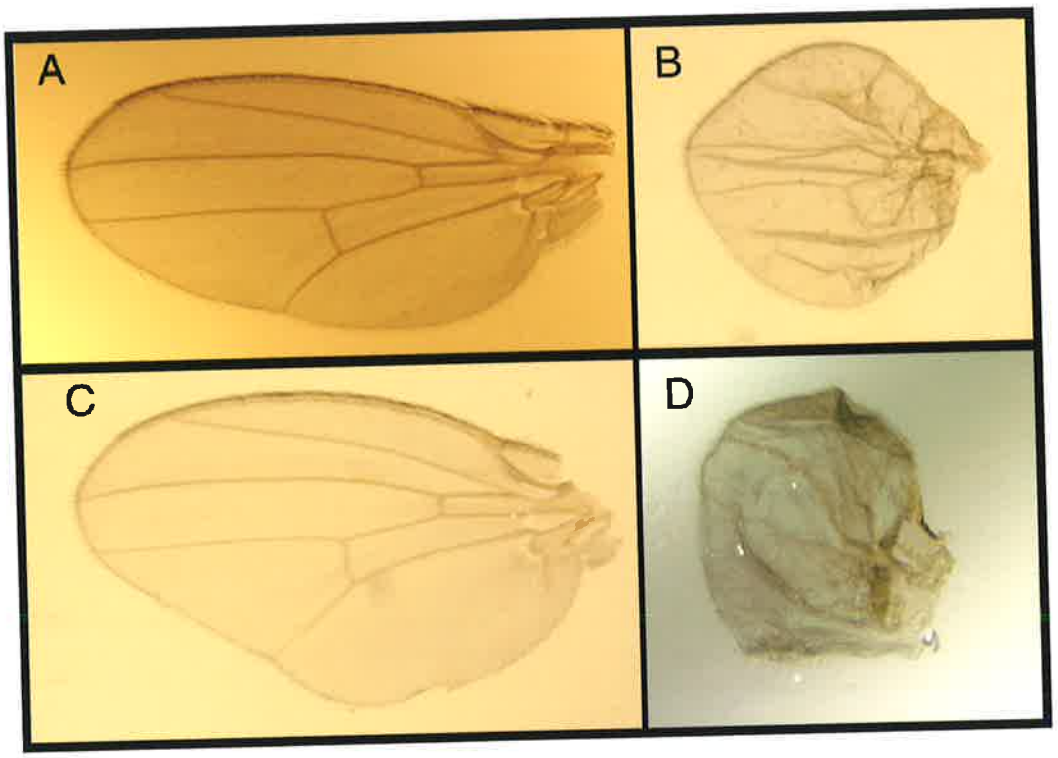
Sequence analysis of the 15,895 bp *dppgen-Am* genomic clone demonstrated that it contains two exons, exon II and exon III, corresponding to the entire coding region of DPP-*Am*, plus the 3′UTR. These are separated by a 923 bp intron, designated intron II (see Figure 3.1A). The position of the splice sites of intron II are conserved throughout the Metazoa (see Figure 3.1B). However, analysis of the length of comparable introns in both *D. melanogaster* and vertebrate *dpp/bmp4* loci reveals

Figure 3.9. Ectopic expression of *dpp* in the *D. melanogaster* wing leads to overgrowth of the wing.

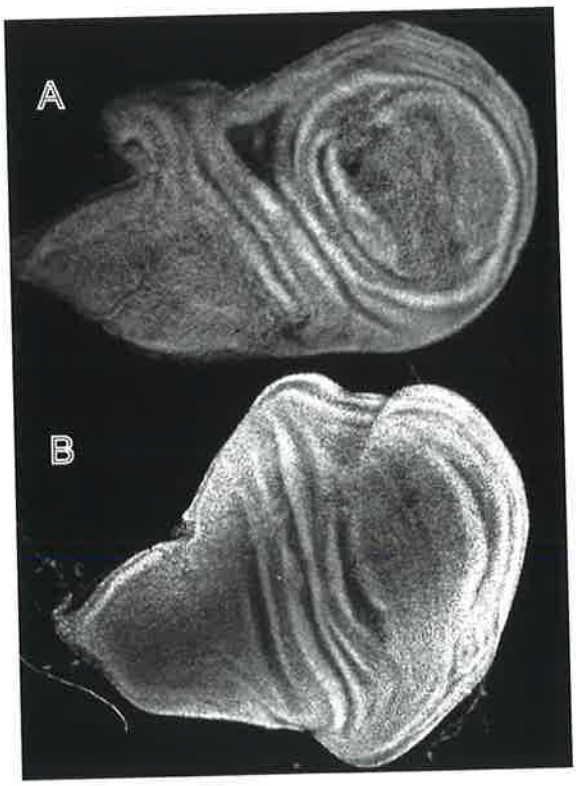
A: Wings from flies ectopically expressing *dpp*: A: a wild-type wing. Wings shown in B, C and D are ectopically expressing *dpp* under the control of the *vestigial* promoter (*vg::*GAL4-UAS::*dpp*). Wings from flies ectopically expressing UAS::*dpp-Dm* are blistered, suggestive of overgrowth (B). This severe phenotype can also be seen in wings from flies ectopically expressing UAS::*dpp-Am*-T3 (D). Wings from flies ectopically expressing UAS::*dpp-Am*-T2 have a milder phenotype, as evidenced by a size increase in the posterior compartment of the wing (C).

B: *D. melanogaster* wing discs stained with Hoechst: B: a wing disc that is ectopically expressing *vg::*GAL4-UAS::*dpp-Am*-T3. This is significantly larger in size than a wild-type wing disc (A) with respect to the anterior/posterior axis, demonstrating overgrowth of the disc.

A



B



that the *D. melanogaster dpp* intron is significantly greater in length than the coral and vertebrate introns (see Figure 3.1B).

The putative exon I is predicted to be greater than 10 kb from exon II. A precedent for this large intron size can be seen in *D. melanogaster,* where the most 5′ exon, designated exon1a, is greater than 20 kb upstream of exon II (St. Johnston *et al.*, 1990). Furthermore, at the mouse *bmp4* genomic loci, the most 5′ exon is greater than 16 kb upstream of exon III (Kurihara *et al.,* 1993).

There are 5 known *D. melanogaster dpp* transcripts (St. Johnston *et al.*, 1990). Each transcript has its own temporal and spatial pattern of expression, yet all transcripts encode an identical polypeptide consisting of three exons, only differing in their most 5′ untranslated exon. Similarly, several *bmp4* transcripts have been identified in vertebrates (for example, Feng *et al.*, 1995). These, too, only differ in their 5′ untranslated exons. Consistent with this, northern analysis of *dpp-Am* shows the presence of a second *dpp* transcript (Hayward *et al.*, submitted; see Figure 1.11). It is thus highly probable that this second transcript only differs from the isolated Dpp-*Am* cDNA by the nature of its 5′ exon. Furthermore, this alternative 5′ exon is likely to only contain 5′ untranslated sequence. Attempts were made to isolate the second *dpp* transcript, although all unfortunately proved unsuccessful. It would thus be interesting to isolate this transcript and identify its putative alternative exon I to see if its position has remained conserved during evolution. Further analysis of both *dpp-Am* and this alternative transcript needs to be carried out, focussing on the detailed characterisation of their promoters and enhancers.

3.4.2 *A. millepora* DPP is a functional homolog of *D. melanogaster* DPP

The data of Hayward *et al.*, submitted, and of this chapter, establishes that DPP is evolutionarily ancient, antedating the separation of the Cnidaria from the rest of the Metazoa.

The question of a conserved ancestral mechanism for the generation of bilaterality using DPP/BMP2/4 signalling is an interesting, important and unresolved one. Molecular, morphological and genetic evidence from a small number of vertebrate and invertebrate model organisms argues strongly for a conserved dorsal/ventral patterning mechanism involving DPP/BMP2/4 signalling. However, the DPP activity gradient is established in very different ways in arthropods and chordates and the

DPP/BMP2/4 pathway is apparently absent or highly diverged in *C. elegans*, suggesting that such signalling may not be essential for bilaterality. The ancestral nature of the role of DPP/BMP2/4 in dorsal/ventral axis formation therefore remains unclear.

The biological activity of DPP-*Am* in *D. melanogaster* implies substantial conservation of the DPP/BMP2/4 signalling system between *A. millepora* and *D. melanogaster*. Specifically, DPP-*Am* is presumably correctly processed in *D. melanogaster,* and must interact with *D. melanogaster* DPP-specific receptors *in vivo,* to generate the same phenotypes as those generated by ectopic *D. melanogaster dpp* expression. Several functions for the highly conserved DPP-*Am* can be speculated. Firstly, Dpp-*Am* could have a role in oral/aboral axis specification, reminiscent of the role of DPP/BMP2/4 in dorsal/ventral axis patterning of higher metazoans. Secondly, whilst the distinction between radiate and bilateral animals is convenient, it has sometimes been regarded as an oversimplification (Willmer, 1990). This leaves open the question of whether the presence of *dpp-Am* in the coral is evidence for a transient second axis. A further possibility is that this sub-class of TGF-β molecules had non-axis specifying ancestral roles.

During the progression of this study, expression analysis of *dpp-Am* was performed in *A. millepora* embryos (Hayward *et al.*, submitted). *In situ* hybridisation demonstrated that mRNA is initially detected at scattered positions in the ectoderm, predominantly on the originally higher side of the inflating embryo (Figure 3.10C). This expression then becomes more intense and solid in one quadrant of the surface ectoderm, adjacent to the blastopore (Figure 3.10D-F). Although the localised expression of *dpp-Am* is typical of the expression pattern seen in genes with axis-specifying roles, it need not necessarily be associated with axis formation. In other animals, localised expression of DPP/BMP2/4 can be attributed to the partitioning of cells to adopt different fates. For example, in sea urchin embryos BMP2/4 determines the position of the boundary between ectoderm and endoderm, and in vertebrates, as well as in invertebrates, DPP/BMP2/4 activity distinguishes non-neurogenic from neurogenic ectoderm (*X. laevis*: Wilson and Hemmati-Brivanlou, 1995, Sasai *et al.*, 1995; *D. melanogaster*: Ferguson and Anderson, 1992a; Sea urchin: Angerer *et al.*, 2000). DPP-*Am* may perform such roles in *A. millepora*.

Unfortunately, these findings still leave the role of DPP in axis specification of *A. millepora* embryos uncertain. It is possible that DPP/BMP2/4 ligands were already

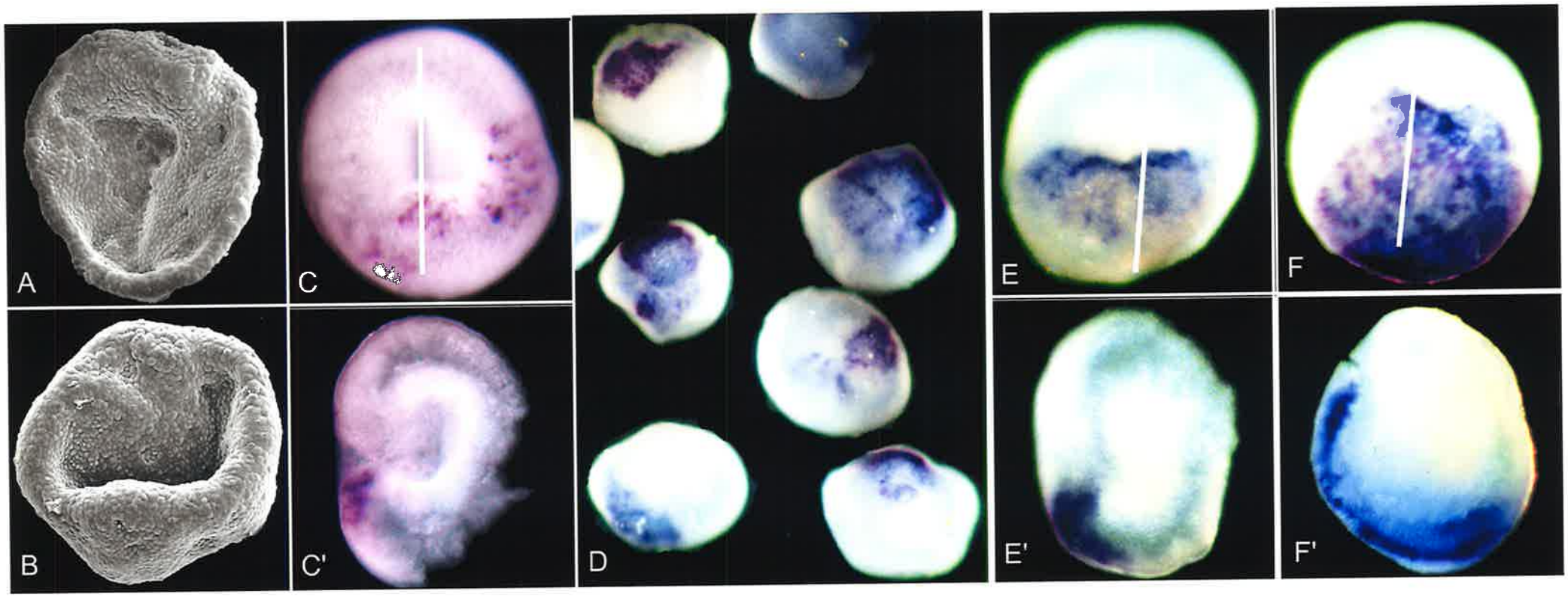
playing a role in axis formation in a putative common ancestor of *A. millepora* and the Bilateria. Alternatively, this family of ligands may have evolved to play one of the non-axis forming roles of DPP/BMP2/4 prior to the diploblast-triploblast divergence. If the latter is true, the localised expression of DPP-*Am* in the region of the blastopore provides a basis on which a DPP/BMP2/4-dependent second axis could have evolved from an ancestral radial animal.

In conclusion, this chapter characterises the *dpp-Am* genomic locus, and illustrates the conservation of its organisation throughout the Metazoa. Further, this chapter demonstrates that this conservation extends to the protein level and provides evidence for coral DPP being a functional homolog of *D. melanogaster* DPP, such that DPP-*Am* can mimic DPP-*Dm* activity in the fly. These results indicate that DPP signalling predates both the overt bilaterality of triploblastic organisms, and limb development, although it is unclear whether the signalling was originally for axis specification or some other role.

The functional conservation of DPP-*Am* predicted the conservation of a functional DPP signalling pathway in *A. millepora*. Chapter 4 details the identification and analysis of both a DPP/BMP2/4-specific receptor, and two DPP/BMP2/4-specific intracellular components of the pathway.

Figure 3.10. *dpp-Am* expression during *A. millepora* development.

A. and B. show scanning electron micrographs of *A. millepora* embryos at the prawn chip stage of development. During this stage no *dpp-Am* expression is detected (data not shown). During the fat donut stage of development *dpp-Am* expression is initially patchy and can be seen within a domain associated with the blastopore (C). A cross-section of the embryo shown in C (C') reveals that this *dpp-Am* expression is limited to the ectoderm. The white line on the embryo indicates the plane of section. D shows a random collection of stained embryos all fixed at a single point in time during this fat donut stage. There is considerable variability in the extent of staining, but in all cases the staining is associated with the blastopore. E and F show a higher magnification view of two embryos from the batch shown in D. E' and F' are cross-sections of the embryos shown in E and F, respectively (white line on embryos in E and F indicates the planes of section). Staining of these embryos confirms that *dpp-Am* transcripts are restricted to the ectoderm. Data courtesy of Dr Eldon Ball, ANU, Canberra.


A
B
C
C'
D
E
E'
F
F'

4. Isolation of components of the DPP signalling pathway in *A. millepora*

4.1 Introduction

The isolation of such a highly conserved *A. millepora* DPP homolog implied the presence of a functional DPP signalling pathway in the coral. To analyse this, and to see whether this signalling pathway, in addition to the DPP ligand, is highly conserved throughout metazoans, it was decided to isolate and characterise coral homologs of proteins known to act downstream of the DPP signal. This chapter describes the identification of an *A. millepora* DPP/BMP2/4-specific TGF-β serine/threonine kinase receptor, along with two intracellular DPP/BMP2/4-specific R-Smads. Furthermore, it details the comparative sequence analyses of these coral signalling components with their higher metazoan homologs.

Comparisons between the expression patterns of these intracellular signalling proteins with those of *dpp-Am* may provide support for their role in DPP signalling. In addition, it may provide greater insight into the function played by the DPP signalling pathway during coral development. Thus, this chapter also details the temporal expression patterns of both the DPP/BMP2/4-specific TGF-β receptor, and the R-Smads, and describes the attempts made to analyse the spatial expression pattern of one of these R-Smad genes.

4.2 Characterisation of Bmpr1-*Am*, an *A. millepora* DPP/ BMP2/4-specific type I receptor

4.2.1 Isolation of *bmpr1-Am*

Initially, degenerate PCR was employed to amplify sequences corresponding to a region of a coral type I receptor. Briefly, primers were designed from conserved residues within kinase domains of different TGF-β type I receptors (see Section 2.1.12.3), and these were used to amplify DNA from an *A. millepora* Prawn Chip cDNA library (see Section 2.1.11). Two rounds of PCR amplification (see Section 2.2.10.2) generated the expected 177 bp PCR product, which was subsequently cloned into the pGEM-T® Easy vector (see Sections 2.2.5-2.2.7 and 2.1.10.1). Sequence analysis of this construct, using the T7 primer (see Section 2.1.12.1), confirmed the presence of an open reading frame corresponding to part of a TGF-β type I receptor. The PCR product was therefore used as a probe to screen 50,000

plaques of the *A. millepora* Prawn Chip cDNA library (see Section 2.2.16). The resulting five positive plaques were excised from their λZAP vector (see Sections 2.2.17.1). Sequence analysis, using the T7 primer, confirmed that they all originated from the same mRNA species. Subsequently, the complete DNA sequence of the clone with the longest cDNA insert was analysed via a stepwise approach in that new primers were designed from each sequencing result (see Section 2.1.12.6). The 3,191 bp clone, referred to as *bmpr1-Am*, contains a complete open reading frame plus 5′ and 3′ untranslated sequences. Predicted translation start and stop sites are positioned at bases 264 and 1856, respectively, giving the predicted protein a length of 530 amino acids and a predicted molecular mass of 60 kDa. The cDNA sequence corresponding to the open reading frame plus its deduced amino acid sequence is shown in Figure 4.1. The complete sequence of the cDNA is detailed in Appendix A.1.

Bmpr1-*Am* displays the characteristic domain structure of a TGF-β type I receptor (Massagué, 1992; see Figure 1.4A and 4.1). A hydropathy plot (DNAstar; DNAstar, Inc, version 5.1) revealed a hydrophobic section between residues 1 and 24 and between residues 133 and 160. These two regions are believed to represent the N-terminal signal sequence and the transmembrane domain, respectively. Separating these two sections is the extracellular domain that contains a cysteine box [CCX_4CN] (see Section 1.6.3.2), a sequence motif present in all TGF-β receptors. In addition, preceding this cysteine box is a distinctive arrangement of cysteine residues [$CXCX_4CX_8CX_5CX_5GCX_{18}CX_{10}$] characteristic of type I receptors (Massagué, 1992). The remaining portion of the polypeptide is predicted to be intracellular. It possesses a GS domain, typically found adjacent to the transmembrane region of type I receptors (see Figure 1.4), followed by sequences indicative of a serine/threonine kinase domain (Hanks and Quinn, 1991). There is no evidence of an extended C-terminal tail characteristic of type II receptors (Massagué, 1992).

4.2.2 Comparative analysis of Bmpr1-*Am*

Comparison of Bmpr1-*Am* with the databases confirmed it to be a DPP/BMP2/4-specific receptor (see Section 2.2.33.1). Alignment of the amino acid sequence with related type I receptor sequences showed the expected low level of identity in the extracellular ligand binding domain and much higher conservation in the intracellular kinase domain. Figure 4.2 shows an alignment of the kinase domains of a representative range of DPP/BMP2/4-specific type I receptors with the Bmpr1-*Am* polypeptide. Within the kinase domain, Bmpr1-*Am* is most similar to the *X. laevis*

Figure 4.1. Nucleotide sequence of the coding region of *bmpr1-Am* along with the deduced amino acid sequence.

The putative hydrophobic signal sequence and transmembrane domain are indicated by a blue underline, the cysteine box is indicated by a red box, as are the additional extracellular cysteine residues. The borders of the kinase domain are represented by purple arrows. The GS box, in the predicted intracellular domain, is indicated by the green underline.

```
                                M  A  K  P  T  M  A  S  V  S  L  L     12
                                ATGGCAAAACCGACTATGGCAAGTGTGTCCCTGCTTC
H  L  V  V  L  L  C  F  L  A  H  G  T  K  S  I  R  C  K  C     32
ACTTGGTAGTTCTTCTCTGCTTTCTGGCTCATGGTACTAAGAGCATACGATGCAAGTGTA
S  T  H  N  C  P  G  D  H  I  N  E  T  C  T  T  E  G  H  C     52
GCACCCACAATTGCCCAGGGGACCACATCAATGAAACGTGTACAACAGAAGGTCACTGTT
Y  K  K  V  E  Q  S  E  E  D  G  L  E  Y  V  T  Y  G  C  L     72
ACAAGAAAGTGGAACAAAGTGAAGAGGATGGTCTAGAGTATGNNACATATGGCTGCCTTC
P  P  E  E  Q  T  T  M  Q  C  K  T  P  N  H  I  H  T  R  L     92
CTCCTGAAGAGCAAACCACAATGCAGNGCAAAACACCAAATCACATTCACACCAGACTGC
L  S  I  E  C  C  S  K  D  L  C  N  D  V  L  Q  P  K  L  P    112
TCTCTATAGAATGTTGTANTAAAGACCTGTGCAACGATGTTTTACAACCTAAGCTACCGA
T  T  A  P  P  T  T  I  T  T  V  E  E  E  T  E  E  A  V  T    132
CAACAGCACCCCCCACAACAATTACAACGGTTGAAGAGGAGACAGAAGAAGCTGTCACAG
E  Q  Y  S  I  L  F  I  S  A  G  V  C  V  A  V  F  V  I  F    152
AACAATATTCTATCCTCTTCATTAGCGCAGGTGTCTGTGTAGCAGTCTTTGTGATTTTCC
L  G  V  L  C  C  R  L  R  A  T  R  S  R  L  P  F  P  F  E    172
TTGGAGTCCTTTGTTGCCGATTGAGAGCGACACGCAGCAGACTTNCCTTTCCCTTTGAAG
V  E  K  Y  G  S  P  Y  M  S  S  G  E  T  L  K  D  M  L  D    192
TGGAGAAGTATGGCAGCCCTTATATGTCTTCAGGGGAAACACTCAAAGACATGCTGGATC
Q  S  S  G  S  G  S  G  L  P  L  L  V  Q  R  T  I  A  K  Q    212
AAAGTTCTGGGAGTGGCTCAGGATTACCACTGCTGGTTCAAAGAACCATTGCTAAGCAAG
V  T  L  V  R  S  V  G  K  G  R  Y  G  E  V  W  Q  A  R  W    232
TGACGTTGGTAAGAAGTGTGGGTAAAGGCAGATATGGTGAAGTGTGGCAAGCAAGATGGC
R  G  E  D  V  A  V  K  I  F  L  S  H  C  E  S  S  W  Q  R    252
GAGGAGAGGACGTGGCTGTCAAAATATTCCTGTCACATTGTGAATCCTCATGGCAGAGAG
E  T  E  I  Y  Q  T  V  L  L  R  H  E  S  I  L  G  F  I  A    272
AAACTGAGATCTACCAGACCGTTTTACTGCGGCATGAGAGCATTCTAGGCTTCATAGCAT
S  D  I  I  G  S  N  Q  V  T  Q  M  Y  L  I  T  D  Y  H  P    292
CAGACATTATTGGAAGCAATCAAGTGACACAGATGTATCTCATAACAGATTATCATCCTT
Y  G  S  L  Y  D  F  L  R  C  H  C  L  N  K  K  T  M  I  R    312
ATGGATCATTGTATGATTTCTTACGATGCCACTGCCTCAACAAGAAGACTATGATAAGGC
L  V  L  S  A  S  A  G  L  T  H  L  H  T  E  I  Q  G  T  K    332
TTGTGTTGTCTGCATCAGCAGGCTTGACGCATCTTCATACTGAAATCCAGGGGACAAAAG
G  K  P  P  I  A  H  R  D  M  K  S  K  N  I  L  V  K  E  N    352
GAAAGCCTCCTATAGCTCATCGTGACATGAAAAGCAAGAACATCCTTGTCAAAGAGAACT
L  T  C  C  I  A  D  F  G  L  A  V  K  Y  S  P  E  T  E  E    372
TGACCTGCTGTATAGCAGATTTTGGACTTGCAGTGAAGTACTCGCCAGAAACTGAAGAGG
V  D  I  K  P  D  T  R  V  G  T  R  R  Y  M  A  P  E  V  L    392
TAGACATCAAGCCAGACACAAGAGTGGGAACACGGCGATACATGGCCCCTGAAGTTCTTG
D  N  A  L  D  S  R  N  F  A  A  F  K  M  A  D  M  Y  S  F    412
ACAATGCGTTGGATTCAAGGAACTTTGCTGCTTTTAAGATGGCAGATATGTACTCATTTG
G  L  V  L  W  E  I  A  R  R  C  F  T  D  E  T  G  L  C  E    432
GATTAGTGTTATGGGAGATTGCCCGAAGGTGTTTTACGGATGAAACTGGACTGTGTGAGG
E  Y  Q  I  P  Y  Y  D  M  L  P  G  D  P  S  F  D  E  V  K    452
AGTACCAGATTCCTTACTACGATATGCTTCCTGGAGATCCTTCTTTTGATGAAGTCAAAA
R  V  V  L  T  D  K  R  R  P  S  V  P  N  R  W  Y  R  D  E    472
GAGTTGTGTTGACAGACAAGAGAAGACCCTCAGTGCCAAATAGATGGTACAGAGATGAGT
C  L  Q  T  M  A  K  L  M  T  E  C  W  A  Q  H  P  A  A  R    492
GTCTCCAAACTATGGCCAAGCTGATGACAGAGTGTTGGGCACAACACCCTGCAGCCCGTC
L  T  A  L  R  V  Q  K  T  L  S  K  L  K  K  S  M  D  F  I    512
TGACAGCCTTGAGAGTTCAGAAAACTTTAAGCAAACTCAAGAAGTCAATGGATTTCATAG
D  Q  P  Y  D  A  D  N  D  S  P  R  T  S  V  T  T  A  *
ACCAACCATATGATGCAGACAATGATAGCCCCAGGACAAGTGTCACCACAGCCTAA
```

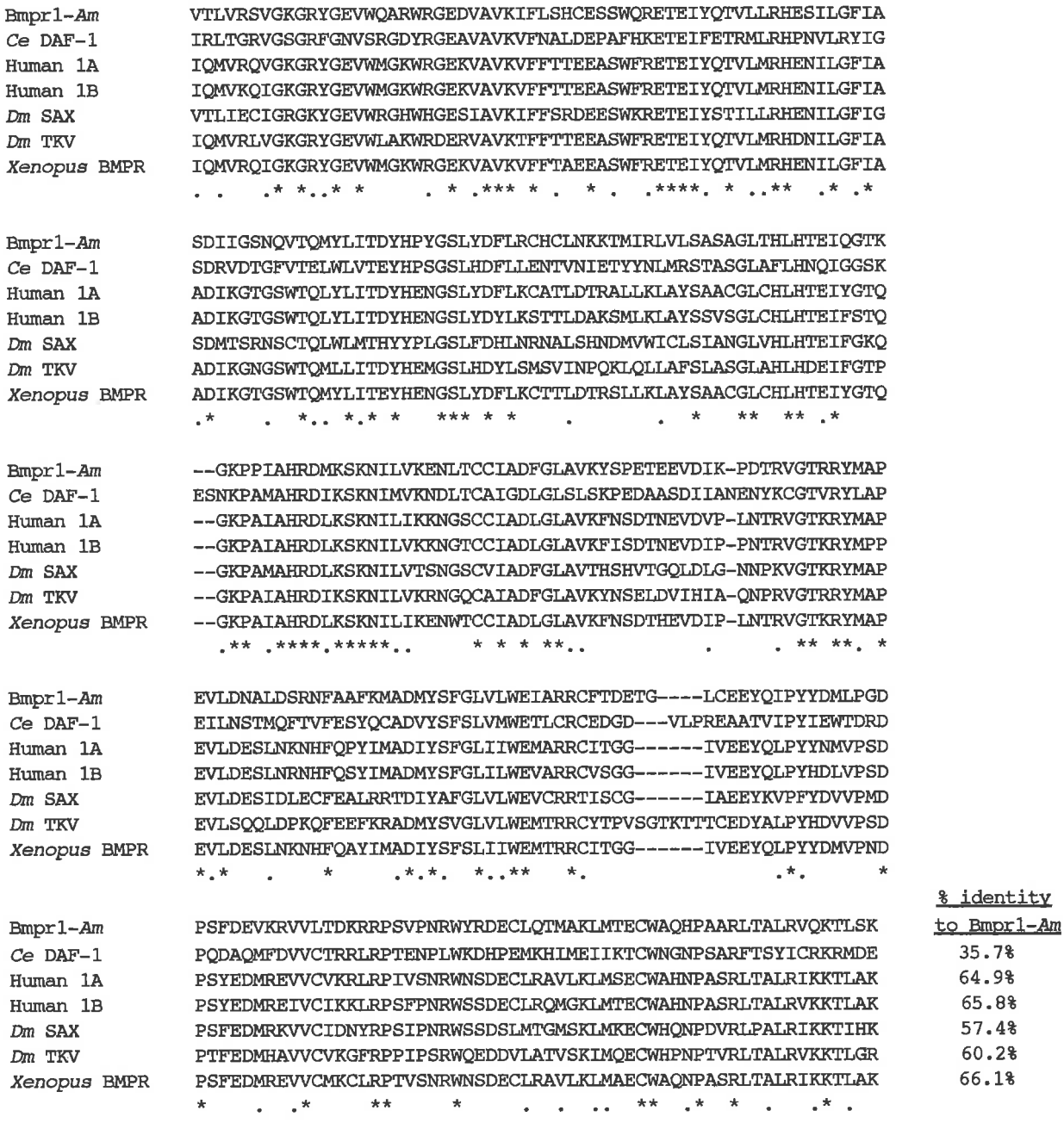


Figure 4.2. Sequence alignment of the predicted kinase domain of Bmpr1-*Am* with the kinase domains of its homologs from other organisms.

Sequences aligned using CLUSTALW. Gaps were introduced to maximise alignment and are shown by dashes. *= identical residues; .= conserved residues. *Dm* = *D. melanogaster*; *Ce* = *C. elegans*.

Bmpr1a receptor (66.1% identity) while identity to *D. melanogaster* TKV and SAX is somewhat lower (60.2% and 57.4% identity, respectively).

Figure 4.3 summarises phylogenetic analyses based on a more extensive range of related molecules, including type II receptors. Note that only the kinase domain was used in the phylogenetic analyses, as unambiguous alignment is only possible for this part of the protein. When the type II receptors were used to define the out-group, these analyses resolved type I TGF-β receptors into three clades, consistent with published phylogenetic analyses (Newfeld *et al.*, 1999). The Bmpr1-*Am* sequence falls basal to the clade constituting the DPP/BMP2/4 sub-class of type I TGF-β receptor sequences. The position of the Bmpr1-*Am* sequence in the analyses highlights the similarity between both the coral protein and the *D. melanogaster* DPP receptor TKV, and the vertebrate type 1a and 1b proteins (the BMP2/4 receptors).

4.2.3 Temporal expression of *bmpr1-Am*

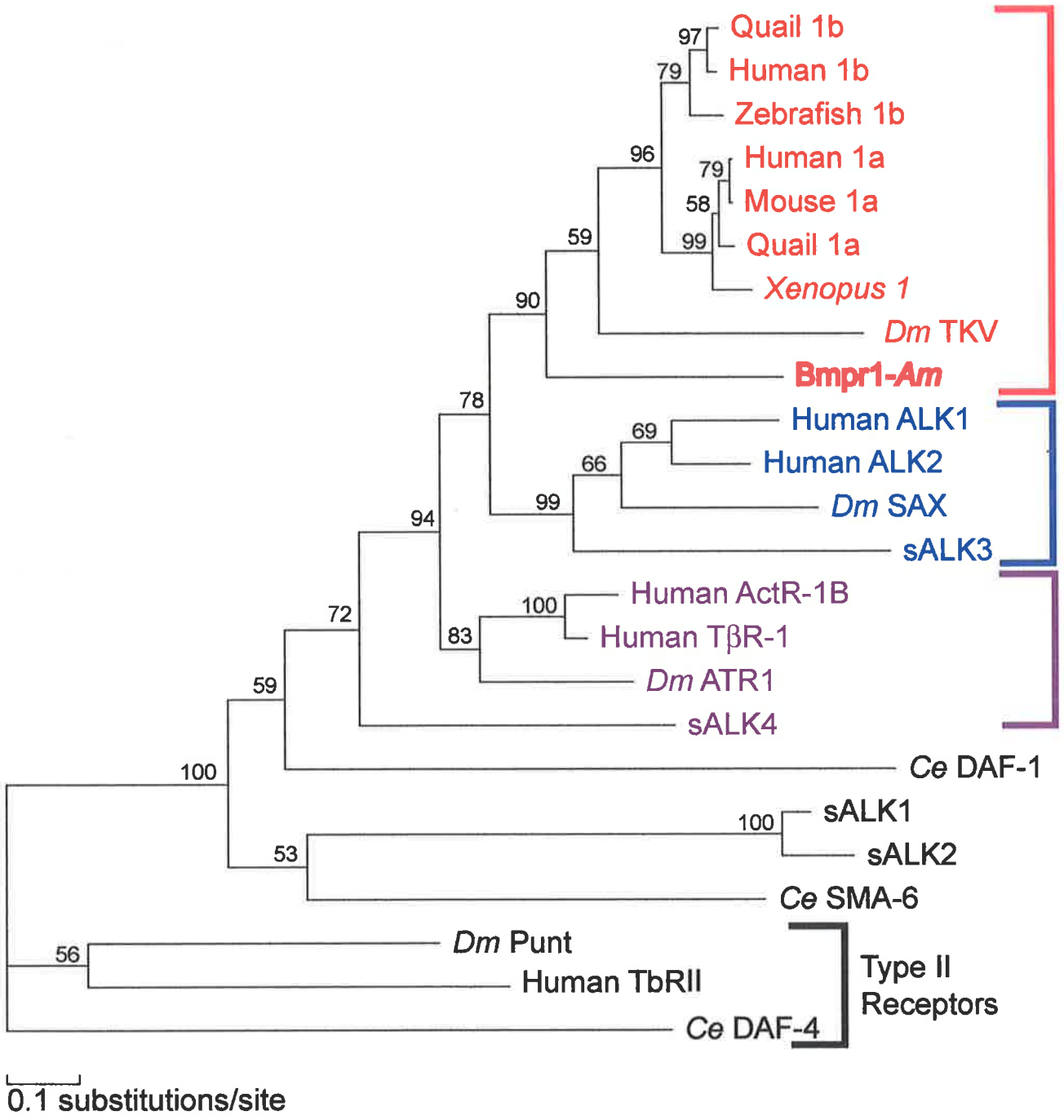
In order to examine the role of *bmpr1-Am* during *A. millepora* development a region of the cDNA was used to probe a northern blot (see Section 2.2.14) which contained mRNA from different stages of *A. millepora* development. Specifically, the Bmpr1-*Am* cDNA was digested with the restriction enzymes *Bgl*II and *Eco*RI, which cut at a single site in the cDNA and the vector multiple cloning site, respectively. This yielded a DNA fragment that contained the 5′-most 1091 bp of the cDNA, corresponding to the 5′UTR and the initial 830 bp of coding region sequence. This DNA fragment was used to generate a probe (see Section 2.2.12). Figure 4.4A shows a single species of approximately 3.0 kb hybridising to this cDNA, demonstrating that the original Bmpr1-*Am* cDNA clone corresponds to the full length transcript. The transcript seemed to be present in equal abundance in both the 24-hour-old and 48-hour-old embryos, as well as in embryos at the pre-settlement stage of development. mRNA isolated from both the *A. millepora* egg and adult was degraded, and due to time constraints, the expression pattern of *bmpr1-Am* at these stages of development remains to be elucidated.

4.2.4 Spatial expression of *bmpr1-Am*

To analyse the spatial expression pattern of *bmpr1-Am*, the Bmpr1-*Am* cDNA was linearised with the *Xba*I restriction enzyme, and used as a template to generate a DIG-labelled RNA probe for *in situ* hybridisation on coral embryos (see Sections 2.2.25.1 and 2.2.31). Due to reasons discussed in Section 4.4, this technique was unsuccessful and the spatial expression pattern of *bmpr1-Am* remains to be elucidated.

Figure 4.3. Phylogenetic relationships of the DPP/BMP2/4-specifc type I TGF-β receptors.

The maximum likelihood analysis was conducted on the amino acid alignment of the kinase domains only (Figure 4.1 illustrates the borders of the kinase domain). Numbers in brackets are GenPep identification numbers. The type II receptors (black brackets on the right hand side) *D. melanogaster* Punt (*Dm*Punt, 784876), *C. elegans* DAF-4 (*Ce*DAF-4, 542467), and human TGF-β type II receptor (Human TβRII, 4507469) were used to root the tree. The *A. millepora* Bmpr1 (Bmpr1-*Am*) protein sequence was compared with: BMP type 1b receptors from quail (6164918), human (4502431) and zebrafish (4586516), the type 1a receptors from quail (6164916), human (4757854) and mouse (547779), the *X. laevis* BMP receptor (*X. laevis*1, 2446990), *D. melanogaster* Thickveins (*Dm*TKV, 2133655), Saxophone (*Dm*SAX, 2133654) and ATR1 (*Dm*ATR1, 436960, product of the *baboon* gene), the human TGF-β and activin receptors ALK1 (3915750), ALK2 (462447), TβR-1 (547777) and ActR-1B (547775), *C. elegans* DAF-1 (*Ce*DAF-1, 118230) and SMA-6 (*Ce*SMA-6, 4219016), and the sponge receptors sALK1 (5596340), sALK2 (5596342), sALK3 (5596344) and sALK4 (5596346). The three coloured brackets on the right hand side of the figure indicate the three clades of TGF-β receptors described by Newfeld *et al.* (1999). Numbers above branches indicate the percentage of 2000 bootstrap replicates supporting the topology shown.


Quail 1b
Human 1b
Zebrafish 1b
Human 1a
Mouse 1a
Quail 1a
Xenopus 1
Dm TKV
Bmpr1-Am
Human ALK1
Human ALK2
Dm SAX
sALK3
Human ActR-1B
Human TβR-1
Dm ATR1
sALK4
Ce DAF-1
sALK1
sALK2
Ce SMA-6
Dm Punt
Human TbRII
Ce DAF-4
Type II Receptors
97
79
96
79
58
99
59
90
78
69
66
99
94
100
83
72
59
100
100
53
56
0.1 substitutions/site

4.3 Characterisation of two *A. millepora* DPP/BMP2/4-specific R-Smads

4.3.1 Isolation of *smad1/5a-Am* and *smad1/5b-Am*

Similar to the isolation of the coral receptor (see Section 4.2.1), degenerate PCR was employed to amplify an initial region of a coral Smad gene. Specifically, DNA from the *A. millepora* Prawn Chip cDNA library was PCR amplified using primers that were designed from conserved amino acid residues within the MH1 and MH2 domains of different R-Smads (see Section 2.1.12.3). Two rounds of amplification (see Section 2.2.10.2), each with 30 cycles and using an annealing temperature of 45°C, yielded the expected 1 kb PCR product. After cloning into the pGEM-T® Easy vector (see Sections 2.2.5-2.2.7 and 2.1.10.1), and sequence analysis using the T7 primer (see Section 2.1.12.1), it became clear that this product was heterogeneous, representing two distinct (but related) Smad cDNAs. These were designated a and b and were subsequently used (independently) to screen 50,000 plaques of the *A. millepora* Prawn chip cDNA library (see Section 2.2.16). The resulting positive plaques were excised from their λZAP vector (see Section 2.2.17.1) and sequence analysis was initially conducted using the T7 primer (see Section 2.1.12.1). Further sequence information was determined via a stepwise approach, in that new primers were designed from each sequencing result (see Section 2.1.12.5).

Screening yielded 8 type-a cDNA clones, including clones corresponding to 3 separate transcript lengths (1,910 bp and 2,124 bp and 2,601 bp). Screening with probe b yielded 9 clones, including 2 transcripts of lengths 1.967 bp and 2,557 bp. An entire coding sequence was present in all the transcripts, and for both genes the different size classes of cDNA transcripts differed only in the length of their putative 3′UTRs. The deduced polypeptide sizes of the R-Smad a and b proteins are 444 and 440 amino acids, respectively. The complete cDNA sequences are detailed in Appendicies A.2 and A.3. Typical of R-Smad proteins, both *A. millepora* Smads possess the highly conserved MH1 and MH2 domains linked by a variable, proline-rich linker region (reviewed by Massagué, 1998, see Figure 4.5). Comparison of the *A. millepora* Smad sequences with the databases (see Section 2.2.33.1) indicated that both genes encode proteins that are most closely related to the Smad1/5 group, (i.e., R-Smads specific to the DPP/BMP2/4 signalling pathway) and hence they were designated *smad1/5a-Am* and *smad1/5b-Am*.

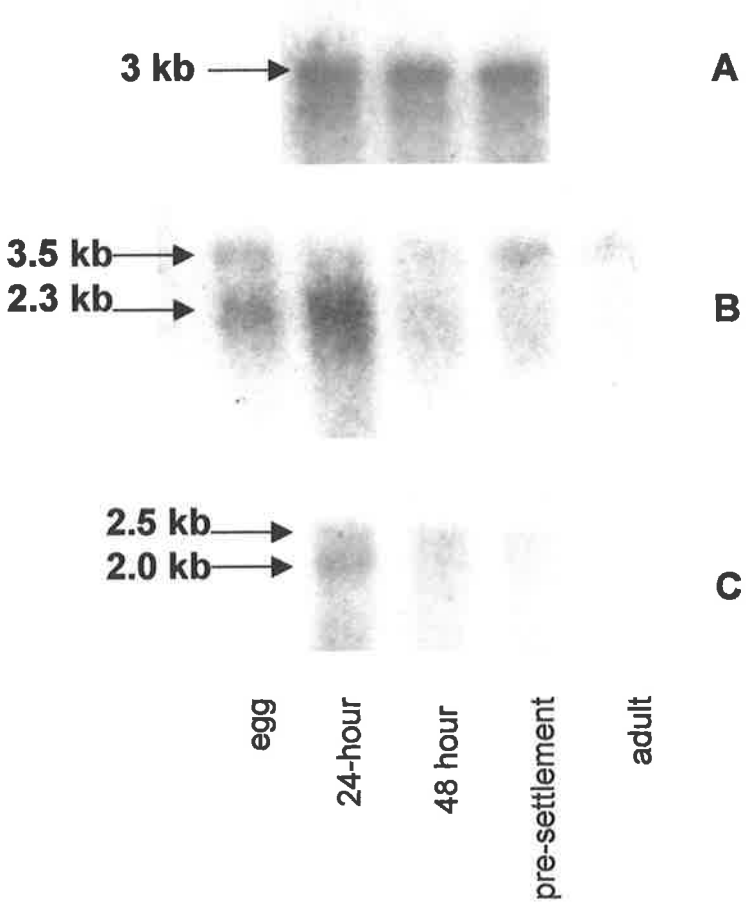
Figure 4.4. Northern hybridisation of the Bmpr1-*Am*, Smad1/5a-*Am* and Smad1/5b-*Am* cDNA to embryonic mRNA.

1μg of embryonic mRNA isolated from different stages of *A. millepora* development was subjected to electrophoresis on three separate 1% agarose gels, and blotted to nylon filters (courtesy of Dave Hayward). The filters were probed with the appropriate ^{32}P-labelled DNA (see text). Sizes of resulting bands were estimated based on migration of marker DNA on the same gel. Each lane is labelled with the developmental stage from which the mRNA was isolated.

A: Northern blot probed with *bmpr1-Am*: A single species of approximately 3 kb can be detected at a similar intensity in 24-hour-old and 48-hour-old embryos, as well as in the embryos at the pre-settlement stage of development.

B: Northern blot probed with *smad1/5a-Am*: Two species can be detected hybridising to the cDNA. These are approximately 2.3 kb and 3.5 kb. The 3.5 kb signal is present during all stages of development analysed. The 2.3 kb species is detected in the egg, with levels increasing dramatically in 24-hour-old embryos.

C: Northern blot probed with *smad1/5b-Am*: Two species can be detected hybridising to the cDNA. These are estimated to be 2.0 kb and 2.5 kb. Both transcripts are present in 24-hour-old embryos. Expression of both transcripts decreases in 48-hour-old embryos, and is weak in embryos at the pre-settlement stage of development.


3 kb
A
3.5 kb
2.3 kb
B
2.5 kb
2.0 kb
C
egg
24-hour
48 hour
pre-settlement
adult

4.3.2 Comparative analysis of the Smad genes

The coral Smad sequences have a high level of identity (approximately 78% identity overall; 89% and 87% in the MH1 and MH2 domains, respectively). Comparisons with Smad sequences of bilateral animals reveals that both the *A. millepora* Smads are most closely related to vertebrate Smad1/5 proteins (see Figure 4.5). In addition, indicative of R-Smads, they contain the characteristic S/TSXS sequence at the C-terminus of the MH2 domain. This is the site of phosphorylation by the type I receptor during ligand-induced activation (Macias-Silva *et al.*, 1996). By contrast, the co- and inhibitory Smads (see Sections 1.6.3.3. and 1.6.3.5) have a maximum of one serine residue in this region (Newfeld *et al.*, 1999). Interestingly, in the case of Smad1/5a-*Am*, this motif is followed by an alanine residue. There are few precedents for this and the functional significance of this residue C-terminal to the phosphorylation site is unknown.

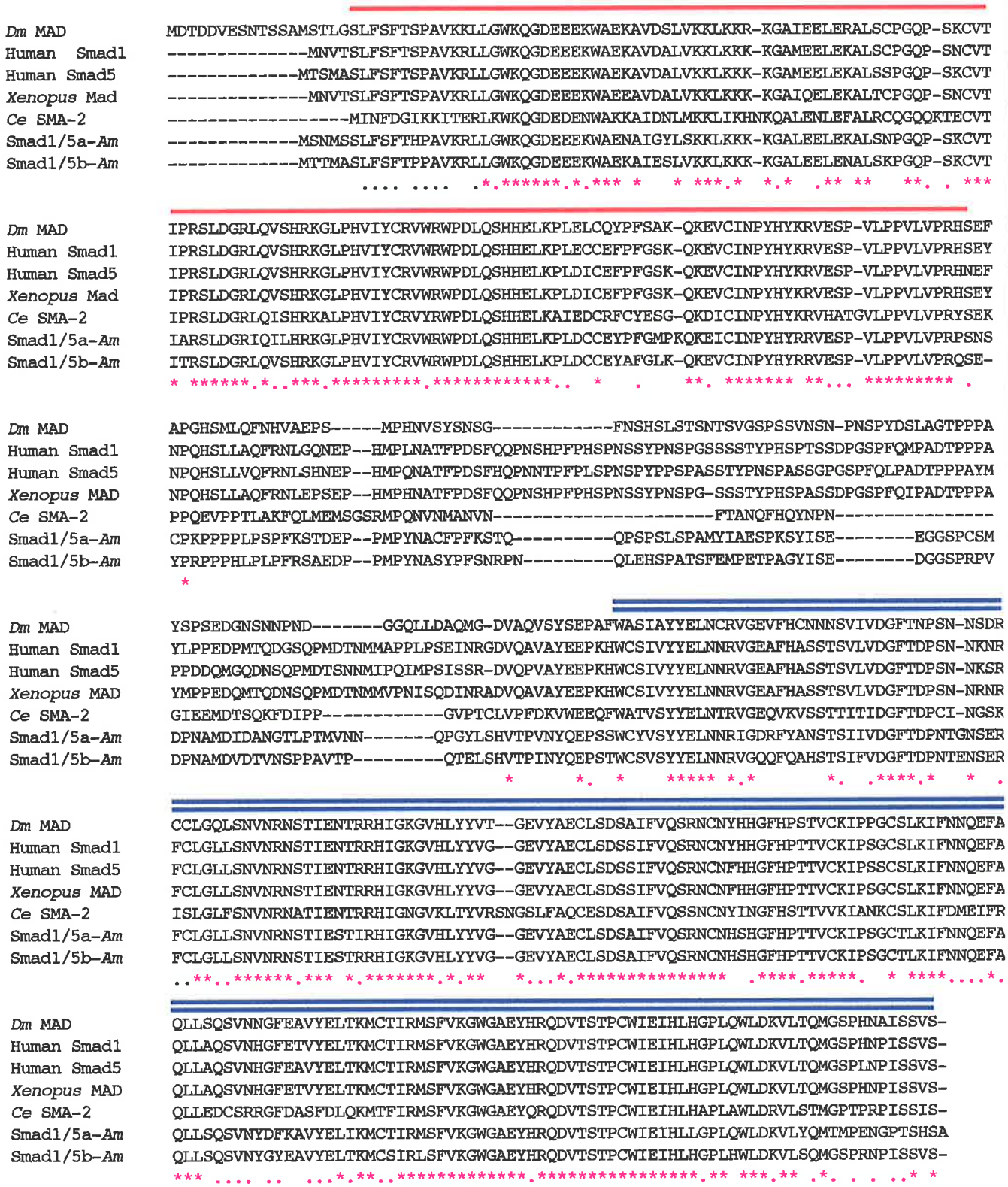
To better understand the relationship between the *A. millepora* Smads and their counterparts in bilateral animals, phylogenetic analyses were conducted using only the MH1 and MH2 domain sequences, as the linker region is sufficiently variable to prevent meaningful alignments between species (see Figure 4.6). In these analyses, the *A. millepora* R-Smads are the basal clade of the DPP/BMP2/4-responsive Smads, and are well resolved from the TGF-β-responsive Smads (Hsmad2 and Hsmad3). The analyses also indicate that Smad1/5a-*Am* and Smad1/5b-*Am* are likely to reflect a cnidarian-specific duplication event (i.e, both are derived from an ancestral cnidarian Smad1/5 gene) and do not correspond to different sub-types from higher animals. Smad1/5b-*Am* has higher levels of identity with Smads in higher animals than does Smad1/5a-*Am*, suggesting that the former may more closely reflect the ancestral state. For example, in the MH2 domain Smad1/5b has 86.1% identity with human Smad1; the corresponding figure for Smad1/5a is 81.2%.

4.3.3 Temporal expression of *smad1/5a-Am* and *smad1/5b-Am*

In order to examine the temporal expression of the two Smad1/5 genes, a northern hybridisation was performed using various stages of *A. millepora* development (see Section 2.2.14). Specifically, the region of cDNA corresponding to the 3′-most sequences (predominantly 3′UTR sequence) of both Smad genes was PCR amplified with specific primers (see Sections 2.1.12.7 and 2.2.10.1). The generated fragments were inserted into the pGEM-T® Easy vector (see Sections 2.2.5-2.2.7 and 2.1.10.1) and sequence analysis confirmed the orientation of the fragments. These constructs were designated pGEM T-*smad1/5a* and pGEM T-*smad1/5b*. The difference in

Figure 4.5. Alignment of the predicted amino acid sequences of Smad1/5a-*Am* and Smad1/5b-*Am* along with their homologs from other organisms.

Sequences aligned using CLUSTALW. Gaps were introduced to maximise alignment and are shown by dashes. *= identical residues; .= conserved residues. The MH1 domain is overlined in red and the MH2 domain is doubly overlined in blue. *Dm = D. melanogaster*; *Ce = C. elegans*.

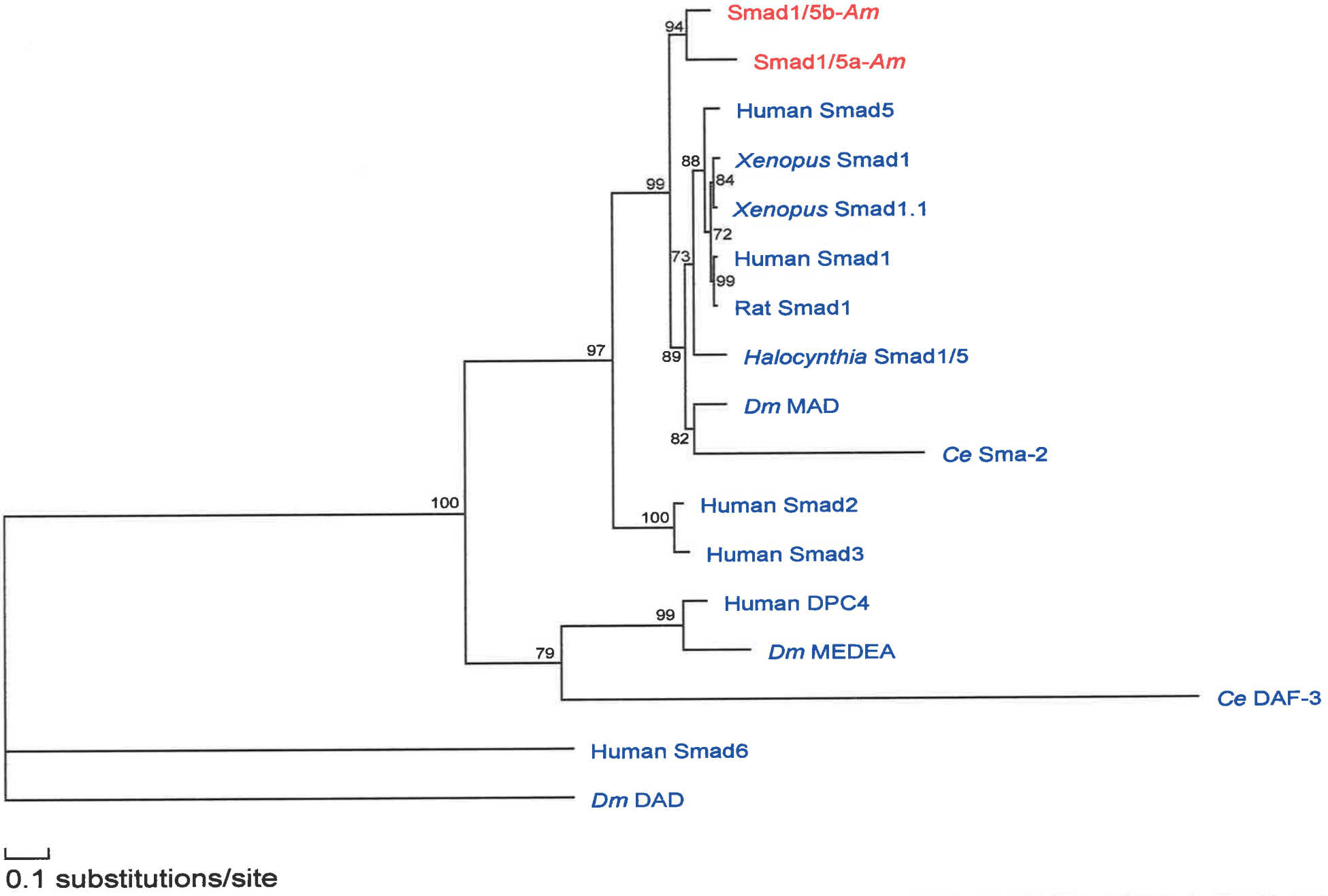


% IDENTITY TO Smad1/5b-*Am*

	Overall	MH1	MH2
Dm MAD	66.4	87.2	82.2
Human Smad1	69.4	88.1	86.1
Human Smad5	70.4	89	86.1
Xenopus Mad	68.8	85.3	86.1
Ce SMA-2	54.9	66.1	63.4
Smad1/5a-*Am*	80.2	86.2	86.6

Figure 4.6. Phylogenetic relationships of the Smads.

The maximum likelihood analysis was conducted on the alignment of the conserved MH1 and MH2 domains only (see Figure 4.5). Numbers in brackets are GenPep identification numbers. In these analyses, the inhibitory Smads *D. melanogaster* DAD (2541864) and human Smad6 (6502523) were used to root the tree. Other sequences compared with the *A. millepora* sequences (Smad1/5a-*Am* and Smad1/5b-*Am*) were the co-Smads *D. melanogaster* MEDEA (3004861), *C. elegans* DAF-3 (2226360) and human DPC4 (4885457), as well as the following: Human Smad1 (5174509), Smad2 (5174511), Smad3 (2351035) and Smad5 (5174515), Rat (*Rattus rattus*) Smad1 (6981172), *X. laevis* Smad1 (1381671) and Smad1.1 (1763545), *Halocynthia roretzi* Smad1/5 (4519908), *D. melanogaster* MAD (1170853) and *C. elegans* Dwarfin SMA-2 (1173452). Numbers above branches indicate the percentage of 2000 bootstrap replicates supporting the topology shown. *Dm* = *D. melanogaster*; *Ce* = *C. elegans*.


Smad1/5b-Am
Smad1/5a-Am
Human Smad5
Xenopus Smad1
Xenopus Smad1.1
Human Smad1
Rat Smad1
Halocynthia Smad1/5
Dm MAD
Ce Sma-2
Human Smad2
Human Smad3
Human DPC4
Dm MEDEA
Ce DAF-3
Human Smad6
Dm DAD
94
88
84
99
72
73
99
97
89
82
100
100
99
79
0.1 substitutions/site

similarity between the 3′ sequences of each gene guaranteed specificity during northern blot analysis. To generate the probes for northern analysis, the Smad sequences were isolated from the vector by restriction digestion with *Eco*RI.

Figure 4.4B shows the results of the northern analysis of *smad1/5a*. Two species, corresponding to sizes of approximately 2.3 kb and 3.5 kb, can be seen hybridising to the cDNA. The 3.5 kb hybridising band is present during all stages of *A. millepora* development analysed, whereas the 2.3 kb band can be seen in the egg, increases sharply in 24-hour-old embryos, and is not expressed, or is expressed at a very low level, during the remainder of development. Although the lower hybridising band was estimated as 2.3 kb, it is broader than expected, and thus is likely to represent all three cDNA transcripts isolated (1,910 bp, 2,124 bp and 2,601 bp). The cDNA corresponding to the 3.5 kb band has not yet been identified. Figure 4.4C shows the results of the northern analysis of *smad1/5b*. It reveals two species, of approximately 2.0 kb and 2.5 kb, hybridising to the cDNA. This is consistent with the sizes of the two *smad1/5b* transcripts isolated (1,967 bp and 2,557 bp), indicating that these cDNAs are full length. Highest levels of both *smad1/5b* transcripts are detected in the 24-hour-old embryo. Expression decreases in 48-hour-old embryos, and is weak in embryos at the pre-settlement stage of development. mRNA isolated from both the *A. millepora* egg and adult was degraded, and due to time constraints, the expression pattern of *smad1/5b* at these stages of development remains to be elucidated.

4.3.4 Spatial expression of *smad1/5a-Am* and Smad1/5a-*Am*

4.3.4.1 Immunohistochemical analysis

The ability to analyse protein distribution is an irreplaceable tool for studying the functions of genes *in vivo*. To this end an antibody was raised against the MH2 domain of Smad1/5a for use during immunohistochemical analysis. This region was chosen as it is highly conserved between *smad1/5a* and *smad1/5b* and should therefore allow cross-reactivity.

Sequence corresponding to the MH2 domain of *smad1/5a* was inserted into the protein expression vector pGEX-2 (see Section 2.1.10.1), to allow Smad1/5a to be expressed as a GST fusion protein. The construct was designated pGEX-*smad*. Specifically, primers designed against sequences flanking the MH2 domain of *smad1/5a* (see Section 2.1.12.4) were used in a PCR reaction (see Section 2.2.10.4) and the PCR product was directionally cloned into the *Bam*HI and *Eco*RI sites of

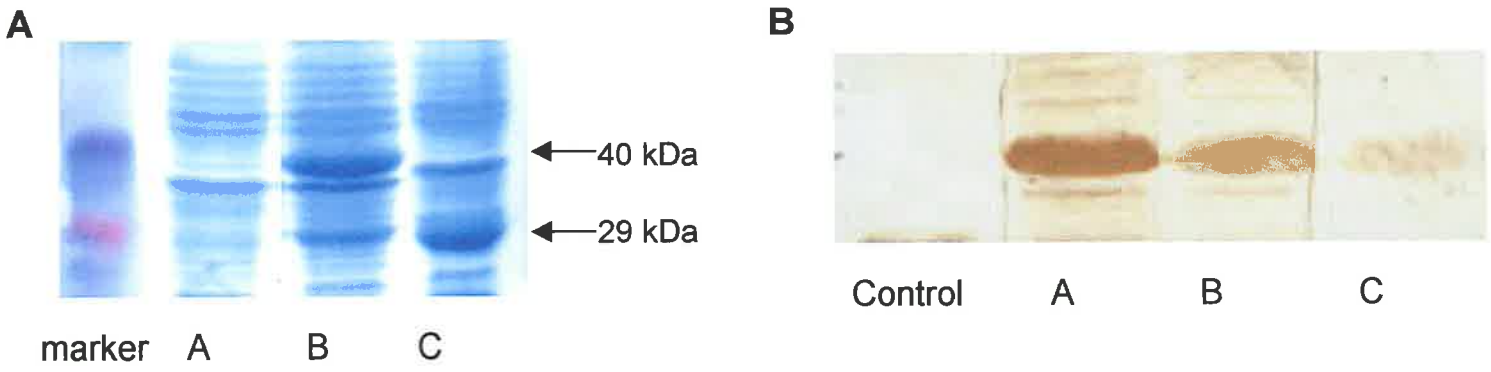


Figure 4.7. pGEX-*smad* was expressed in bacterial cells and used to raise a Smad-specific antibody in rats, termed anti-GST-Smad-*Am*.

A: Expression of pGEX-*smad* in *E. coli*: pGEX-*smad* was expressed in bacterial cells. Total bacterial proteins were separated on an acrylamide gel by electrophoresis, and stained with coomasie. Lane C contains total protein from bacterial cells expressing pGEX alone. There is a band at 29 kDa corresponding to the GST protein. Lane B contains total protein from bacterial cells expressing pGEX-*smad*. There is a specific band at 40 kDa. This band is not present in bacterial cells which contained, but were not induced to express, the pGEX-*smad* construct (lane A).

B: Western immunoblot illustrating the detection of pMAL-*smad* using anti-GST-Smad-*Am*: pMAL-*smad* was expressed in bacterial cells. The bacterial proteins were separated on an acrylamide gel by electrophoresis, and subsequently transferred to a nylon membrane. This membrane was probed with anti-GST-Smad-*Am*. Sizes of resulting bands were estimated based on migration of protein markers on the same gel. A strong band was detected at 52 kDa, which corresponds to the predicted molecular weight of the fusion protein. Concentrations of anti-GST-Smad-*Am* used during the analysis to detect pMAL-*smad*: Lane A = 1/1000, lane B =1/5000, lane C = 1/10 000.

pGEX-2. The fidelity of PCR amplification, and orientation and frame of insertion of the cloned DNA, were confirmed by DNA sequence analysis. The GST fusion protein was expressed in bacterial cells (see Section 2.2.19). To confirm induction, bacterial proteins were separated using SDS-PAGE, and stained with Coomassie Brilliant Blue stain (see Section 2.2.20). This revealed an expected band of approximately 40 kDa, which was absent in non-induced cells (Figure 4.7A). This fusion protein was used to immunise rats (see Section 2.2.22). To determine the specificity of the antiserum, a further construct containing the same *smad1/5a* C-terminal fragment was inserted into the *Eco*RI and *Pst*I sites of the protein expression vector, pMAL (see Section 2.1.10). This system allows Smad1/5a to be expressed in bacterial cells fused to a maltose-binding protein. The construct was designated pMAL-*smad*. Bacterial proteins expressing pMAL-*smad* were analysed via western immunoblots (see Section 2.2.21) using the antiserum as a primary antibody. A specific band at the estimated size of 52 kDa was detected (Figure 4.7B). This confirmed that the antiserum could specifically recognise the Smad1/5a protein when bacterially expressed, so the antibody component of the antiserum was purified by passing it through a Hi-Trap G column. The resulting eluate was designated anti-GST-Smad-*Am*. Immunolocalisations were performed on coral embryos (see Section 2.2.32) using anti-GST-Smad-*Am* as the primary antibody at varying concentrations (1/100, 1/500, and 1/1000). As a negative control several embryos were subjected to the immunolocalisation procedure without primary antibody. Unfortunately the analysis was unsuccessful and no specific staining could be detected in any embryos. The possible reasons for this are discussed in Section 4.4.

4.3.4.2 *In situ* hybridisation

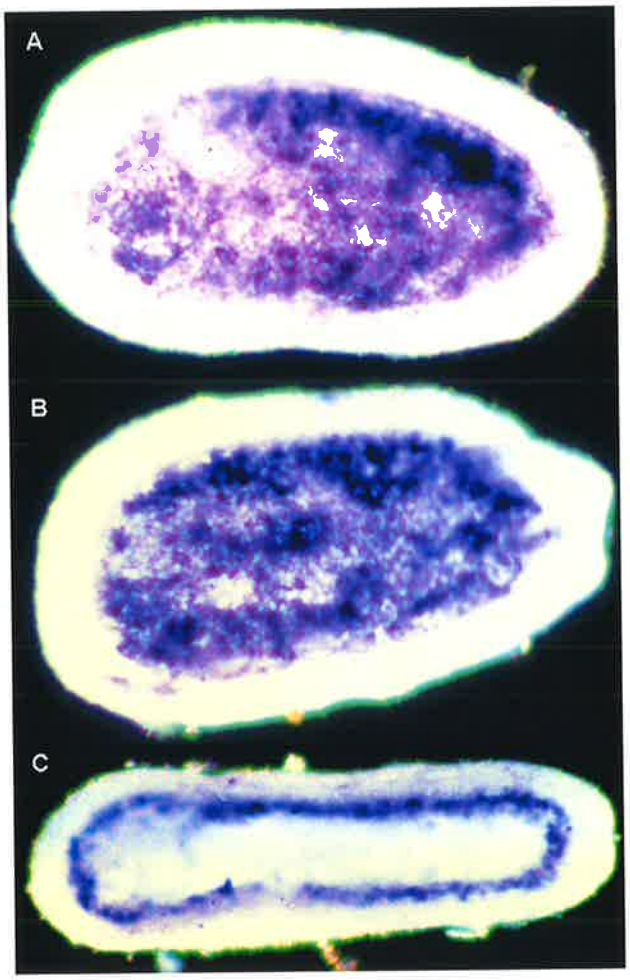
To analyse the spatial expression pattern of *smad1/5a,* pGEM T-*smad1/5a* (see Section 4.3.3) was linearised with *Spe*I and used as a template to generate a DIG-labelled RNA probe for *in situ* hybridisation on coral embryos (see Section 2.2.31). An RNA probe generated from a *D. melanogaster* brain-specific gene was hybridised to the coral embryos and larvae to act as a negative control. Figure 4.8 shows the results of this experiment. There is clear endodermal staining in both the pear stage embryos and the planula larva. However, endodermal staining has consistently been demonstrated to be an artefact of similar experiments on coral embryos (see Section 4.4). This problem has not yet been overcome, making these data difficult to interpret and essentially inconclusive.

Figure 4.8. *smad1/5a-Am* expression in the *A. millepora* embryo and planula larva

In situ hybridisation was performed on *A. millepora* embryos and planula larvae, using the 3′UTR of *smad1/5a-Am* as a probe.

(A and B) **Pear stage *A. millepora* embryos**: *smad1/5a* transcripts are detected specifically in the centre of the embryo. This region corresponds to the endoderm.

(C) ***A. millepora* planula larva**: *smad1/5a* transcripts are detected as an inner ring, immediately inside of the external ectoderm. This region corresponds to the endoderm.


A
B
C

4.4 Discussion

This chapter details the isolation and characterisation of an *A. millepora* DPP/BMP2/4-specific type I receptor and two DPP/BMP2/4-specific R-Smads. These data indicate the presence of a receptor serine/threonine kinase-mediated signalling pathway of the DPP/BMP2/4 type in *A. millepora*.

It is highly probable that Bmpr1-*Am* and the *A. millepora* R-Smads mediate the DPP-*Am* signal. In agreement with this, northern analysis demonstrates that the receptor and R-Smad genes are expressed at the same developmental stages as *dpp-Am*. Indeed, expression of both *dpp-Am* and the R-Smads is highest in the 24-hour-old embryo, and reduced at other stages of development (compare Figures 1.11, 4.4B and 4.4 C).

Members of the TGF-β receptor family of proteins have recently been identified in the sponge *Ephydatia fluviatilis* (Suga *et al.*, 1999), so the presence of related signal transduction pathway components in a cnidarian is perhaps not surprising. However, the sponge proteins are divergent members of receptor types responsive to other classes of ligands and contain no DPP/BMP2/4-specific amino acid residues. In contrast to this the *A. millepora* proteins (a DPP/BMP2/4 receptor and two receptor-mediated Smads) are clearly of the DPP/BMP2/4 type (see below). Indeed, the level of identity between each of the *A. millepora* proteins with the corresponding components of the DPP/BMP2/4 pathway in *D. melanogaster* and chordates is unexpected and striking.

Phylogenetic analyses resolve type I receptors into three distinct classes (Newfeld *et al.* 1999). These three receptor classes have distinct R-Smad specificities, determined by the interaction of the L45 loop (Feng and Derynck, 1997; Chen *et al.*, 1998) with two residues in the L3 loop of the R-Smad MH2 domain (Lo *et al.*, 1998). The sponge type I receptors sALK-3 and sALK-4 have L45 sequences that are intermediate between consensus sequences of TGF-β and activin-responsive receptors (Chen and Massagué, 1999), while the receptors sALK-1 and sALK-2 have highly divergent sequences unrelated to all other type I receptors, presumably reflecting sponge-specific derivation. However, in the case of Bmpr1-*Am*, two of the three specificity-conferring residues in the L45 loop are identical to the DPP/BMP2/4 sub-type, and the third is a conservative substitution (Ser for Thr). Further, both Smad1/5a-*Am* and Smad1/5b-*Am* have the two residues that define R-Smads responsive to DPP/BMP2/4-mediated signalling (His403, Asp406 in Smad1/5a-*Am* and His400 and

Asp403 in Smad1/5b-*Am*). Indeed, the L3 loop sequences in the two coral proteins are identical with those of *D. melanogaster* MAD and many of the vertebrate Smad1/5 types. These above results are consistent with the theory that ancestral type I receptors were of the activin/TGF-β-type, and the DPP/BMP2/4 type arose after the Porifera/(Cnidaria + bilateral Metazoa) split.

Interestingly, phylogenetic analyses of the type I receptors indicate that the Bmpr1-*Am* polypeptide has a higher level of identity with the DPP/BMP2/4-specific type I receptors than does any related molecule from *C. elegans* (including the putative DPP receptor, SMA-6) (Krishna *et al.*, 1999). The evolutionary relationship of the *C. elegans* receptors is reflected in the relationship of their ligands; both the SMA-6 ligand DBL-1, and the DAF-1 ligand DAF-7 show relatively low levels of identity with the DPP/BMP2/4 class (for phylogenetic analysis see Newfeld *et al.*, 1999). Further, the *C. elegans* DPP/BMP2/4-specific R-Smad protein also shows surprising divergence. Although maximum likelihood analyses (see Figure 4.6) are consistent with *C. elegans* SMA-2 being homologous with *D. melanogaster* MAD, the branch length indicates that the SMA-2 sequence is highly diverged. Indeed, sequence identities are higher between *A. millepora* and *D. melanogaster*/vertebrate R-Smads than between SMA-2 and the latter. These observations in *C. elegans* are consistent with the DPP pathway having undergone a high degree of secondary modification in the nematode lineage.

Rather than being homologs of distinct vertebrate R-Smad sub-classes, *smad1/5a-Am* and *smad1/5b-Am* reflect a duplication event that post-dated the divergence of cnidarians and higher metazoans. Since this split occurred at least 540 Mya (Grotzinger and Knoll, 1995), it is perhaps not surprising that several gene classes have undergone duplication events in the Cnidaria since that time. There are precedents for the presence of duplicated R-Smads in *A. millepora*. In *Hydra*, the paired-like genes *prdla* and *prdlb* are both likely to have arisen from the precursor of *D. melanogaster aristaless* (Gauchat *et al.*, 1998), and *Cnnos1* and *Cnnos2* are both related to *D. melanogaster nanos* (Mochizuki *et al.*, 2000).

Even so, the significance of the presence of two R-Smads in *A. millepora* is unclear. Since they have identical L3 sequences, both are likely to transmit signals originating from DPP (rather than activin or TGF-β) molecules. It is also unlikely that they have different intrinsic DNA-binding properties, because there are only minor differences in the DNA-binding region, and the residues likely to make contacts (Arg75, Thr77 and

Lys82; Shi *et al.*, 1998) are identical in the two proteins. It is possible, however, that the two R-Smads have differing protein interaction specificities. For example, the differences in the C-terminal region of the MH2 domain, which is involved in oligomerisation (Shi *et al.*, 1997), suggest that the two R-Smads may interact differently with co-Smads. Alternatively, it is plausible that the two Smad1/5-*Am* proteins perform an identical function in the coral but have a differential spatial expression pattern due to complementary degenerative mutations in the regulatory regions of the genes (Force *et al.*, 1999). This would predict that each Smad gene has experienced a loss of expression for different sub-functions by degenerative mutation, and thus the combined action of both Smads is necessary to fulfill the requirements of the ancestral locus. To assess the likelihood of this latter possibility, both Smad genes will need to be analysed by *in situ* hybridisation.

Another difference between the two *A. millepora* R-Smads is evident at the C-terminus of these proteins. The S/TSXS phosphorylation site is at the C-terminus of almost all known R-Smads. Smad1/5a-*Am* has a C-terminal extension of a single alanine residue. The effect of this residue on phosphorylation of the S/TSXS motif, and consequently on the interaction with the co-Smad and nuclear localisation, is unknown. A precedent for an extension is provided by *X. laevis* Smad1.1, which carries a 3 residue (LeuMetAsp) C-terminal extension (Meersseman *et al.*, 1997). This protein was shown to be localised in the nucleus after overexpression in COS cells (Meersseman *et al.*, 1997). Since phosphorylation of the protein is required for its activation and subsequent translocation to the nucleus, this result predicts that the phosphorylation of *X. laevis* Smad1.1 is unaffected by its C-terminal extension.

Although *smad1/5a-Am* transcripts appear to be detected in the endoderm of both pear stage *A. millepora* embryos and the planula larvae, by *in situ* hybridisation, it is highly probable that this is artefactual. Protocols describing whole mount *in situ* and immunohistochemical analysis on coral embryos have only been recently documented, and there are still many stumbling blocks that need to be addressed. The foremost of these is the artefactual endodermal staining seen in the negative controls of many experiments, including the experiment carried out in this thesis. At present, the time taken for the endoderm of the embryos to stain, compared to those of a negative control, is sometimes the only indication of true staining. Unfortunately, even this method of analysis was not a good enough indication to make conclusions about the expression pattern of *smad1/5a-Am*. In spite of this, and unlike the *dpp-Am* expression pattern (see Figure 3.10), no ectodermal staining of *smad1/5a-Am* could

be seen. Although it is likely that this is due to failure of the *in situ* hybridisation procedure, it is possible that it is a true result. There is a precedent for this. In *D. melanogaster*, in some instances, *dpp* is expressed in one germ layer and travels to another to elicit its effect (Frasch, 1995; Maggert *et al.*, 1995). If the function of DPP-*Am* is to signal to endodermal cells, then the signal transduction components may be restricted to the endoderm.

In conclusion, this chapter details the isolation and characterisation of an *A. millepora* DPP/BMP2/4-specific type I receptor and two DPP/BMP2/4-specific R-Smads. These data indicate the presence of a receptor serine/threonine kinase-mediated signalling pathway of the DPP/BMP2/4 type in *A. millepora*. Further expression and functional work will need to be performed to determine what role this signalling system plays during coral development.

5. An *A. millepora* EST analysis

Introduction

As discussed in previous chapters, cnidarians, in particular corals, represent a perfect outgroup for developmental and evolutionary studies of higher metazoans. Despite this, very little molecular developmental studies of coral have been attempted and, as with any newly studied animal, work is hampered by limited experimental knowledge. Only a handful of *A. millepora* genes have been isolated and because of the inability to perform any functional analyses to date, the roles of many of these genes remains speculative. Further, at present very few suitable *A. millepora* marker genes, either spatial or temporal, have been identified. This chapter describes work performed in order to facilitate the identification and analysis of *A. millepora* genes, in particular potential tissue- and developmental stage- specific marker genes, which would provide valuable information regarding gene expression patterns during coral development. This involved a limited Expressed Sequence Tag (EST) analysis.

5.2. *A. millepora* EST analysis

5.2.1 Isolation of ESTs

cDNA inserts isolated from a directionally cloned *A. millepora* pre-settlement cDNA library (96 hours post-fertilisation; see Sections 1.4.1 and 2.1.11) were employed for this EST study. By this stage of development coral embryogenesis is complete and the embryo has developed into a fully differentiated planula larva, with a simple nerve net and cilia. Thus it was hoped that, in contrast to an earlier developmental stage, a more varied array of potentially interesting genes would be expressed in the different cell types of the larva.

As many highly significant developmental genes are expressed at low levels, their transcripts are poorly represented in the total population of mRNA in a cell. For this reason, an equalisation step was employed to increase the proportion of rare cDNAs in the *A. millepora* library, and to decrease the levels of highly expressed sequences. Specifically, LL-SseIA-tagged (see Section 2.1.12.10) T3 and T7 primers (Norm1 and Norm2; see Section 2.1.12.10) were used to amplify the cDNA inserts of the library (see Section 2.2.10.1). These cDNAs were subsequently equalised using limited re-association of denatured strands (see Section 2.2.11), and selectively amplifying the single strands in a PCR reaction, as above, except that the LL-SseIA sequence (see Section 2.1.12.10) was employed as a primer. cDNA insert lengths within the equalised library averaged approximately 900 bp, a similar insert length to that seen in

the original library. To verify the success of the equalisation procedure, the highly abundant *A. millepora* Ub52 gene (Berghammer *et al.*, 1996) was hybridised to equal amounts of both the pre-equalised and equalised cDNA inserts during a southern hybridisation (see Section 2.2.13). Specifically, the *ubi*-CEP-52 construct (see Section 2.1.10.1) was restricted with the *Eco*RI and *Not*I enzymes. This yielded a DNA fragment of 360 bp, corresponding to *ubi-cep52*, which was subsequently used to generate a probe. Less probe hybridised to the equalised library than to the pre-equalised library demonstrating that there is a decreased representation of the highly abundant *A. millepora* Ub52 gene in the former library (Figure 5.1). With the success of the equalisation step confirmed, the cDNA inserts were ligated to the pGEM T® Easy vector and transformed into bacterial cells (see Section 2.2.5). The subsequent bacterial colonies were individually screened, via PCR amplification using the LL-SseIA primer (see Sections 2.1.12.10 and 2.2.10.3), for the presence of a cDNA insert. Only those colonies containing a plasmid with a cDNA insert of greater than 500 bp were selected. This ensured that during later analysis a meaningful length of sequence data could be studied. In total, 3000 colonies were selected, and transferred to 96-well plates for storage at -80°C as glycerol stocks (plates A-Z; a-f).

5.2.2. Sequence analysis of ESTs

As cDNAs originating from an identical mRNA may have different 5′ terminal sequences, sequencing from these termini may have resulted in duplicated ESTs being unrecognised. It was decided, therefore, to analyse DNA sequence from the 3′ termini of the cDNA inserts. The T3 primer sequence (see Section 2.1.12.1) is not present in the pGEM T® Easy vector, but was originally used to amplify the cDNA inserts from the directionally cloned pre-settlement library (see Section 5.2.1). Thus, this primer was employed in the sequencing reaction as it would not only give insert-specific sequence data, but it also ensured that the sequences would correspond to the 3′ termini of the cDNA inserts. Sequence analysis was performed by the Australian Genome Research Facility (AGRF), Brisbane, Queensland.

After removal of additional vector sequence and poor sequence data (see Section 2.2.33.2), 2337 sequence reads remained of sufficient quality to be subjected to further analysis. The ESTs were translated in all six reading frames and subjected to a BLASTx search (Altschul *et al.*, 1997) using the Swiss Prot and SpTrEMBL database at Bionavigator (see Section 2.2.33.2). The results indicated that 1808 of these ESTs matched known sequences in the database; 499 of these hits were shown to be redundancies in the EST collection. Of the remaining 1309 sequences, 800 failed to

Figure 5.1. Southern Hybridisation of the pre-equalised and equalised *A. millepora* library.

An equal amount of DNA was loaded onto the lanes of a 1% agarose gel as follows:

Lane 1: cDNA inserts of the equalised *A. millepora* library

Lane 2: cDNA inserts of the pre-equalised *A. millepora* library

The DNA was subjected to electrophoresis (see Section 2.2.3), and blotted to nylon filters (see Section 2.2.13). The filters were probed with ^{32}P-labelled *A. millepora ub52*. Less probe is seen hybridising to the cDNA inserts from the equalised library than the pre-equalised library, demonstrating a decrease in the abundance of this gene in the equalised library.

show a significant match with known sequences at a significance threshold (E) of 10^{-6}; the remaining 509 sequences were examined to determine putative function on the basis of homology (see Appendix A.4). The sequences were categorised by function and the classification system that was employed essentially divided the identified ESTs into 7 classes (see Figure 5.2). These classes included proteins involved in DNA structure, replication and repair (A; 3.1%), transcription factors (B; 5.1%), proteins involved in RNA-binding and splicing (C; 3.6%), proteins involved in translation (D; 13.8%), house-keeping proteins (E, F, G and H; 43.6%), signalling/regulatory molecules (I; 8.6%) and others (J; 22.2%). Certain categories were further sub-divided. These sub-divisions are displayed in Table 5.1, along with the respective number of ESTs contained within each sub-class.

The most represented proteins were ribosomal proteins (see Table 5.1 and Appendix A.4), with as many as 60 distinct ESTs being identified. As expected, the majority of identified ESTs corresponded to house-keeping proteins, including metabolic proteins, proteins involved in protein degradation and processing, transport proteins, and proteins associated with the cytoskeletal framework (see Table 5.1). In the latter class, homologs of a myosin heavy chain, a human unconventional myosin (myosin xv), tropomyosin, and 2 myosin light chains were identified, in addition to 3 actins, two actin-related proteins, and 4 tubulin proteins. Further, a variety of other cytoskeletal proteins were isolated including a profilin homolog, a collagen homolog, a neural tropomodulin homolog, a β-filamin homolog, a spectrin homolog, a vinculin homolog, a dystrophin-like homolog, a radixin homolog, an espin homolog, 6 kinesin or kinesin-like homologs, one of which was neuronal, and 6 dynein homologs, including both light and heavy chains. Proteins were also isolated that were specifically involved in cell adhesion, including homologs of α- and β-catenin, polycystic kidney disease-associated protein, *D. melanogaster* Furrowed, and laminin. The catenins and polycystic kidney disease-associated protein also function as signalling intermediates. A sequence matching Kinectin which, in mammals, interacts with RhoG(GTP) in a microtubule-dependent manner (Vignal *et al.*, 2001) was represented. Cytoskeletal regulatory proteins identified included a sequence matching a myosin phosphatase regulatory subunit, a pest phosphatase interacting protein homolog, and a neurocalcin homolog. The latter is neural-specific in higher animals, interacting with Calcium (Ca^{2+}) in a Ca^{2+}-dependent manner (see Section 5.4.1). A homolog of homer-2b, another cytoskeletal protein isolated, is also neural-specific in higher metazoans (see Section 5.4.1).

Figure 5.2. Pie chart functionally classifying the ESTs which had significant matches to known proteins.

The 509 ESTs with significant BLASTX matches were classified according to their putative biological function. Functional categories are as follows:

A. Proteins involved in DNA structure, replication and repair

B. Transcription factors and associated proteins

C. RNA-binding proteins and proteins involved in splicing

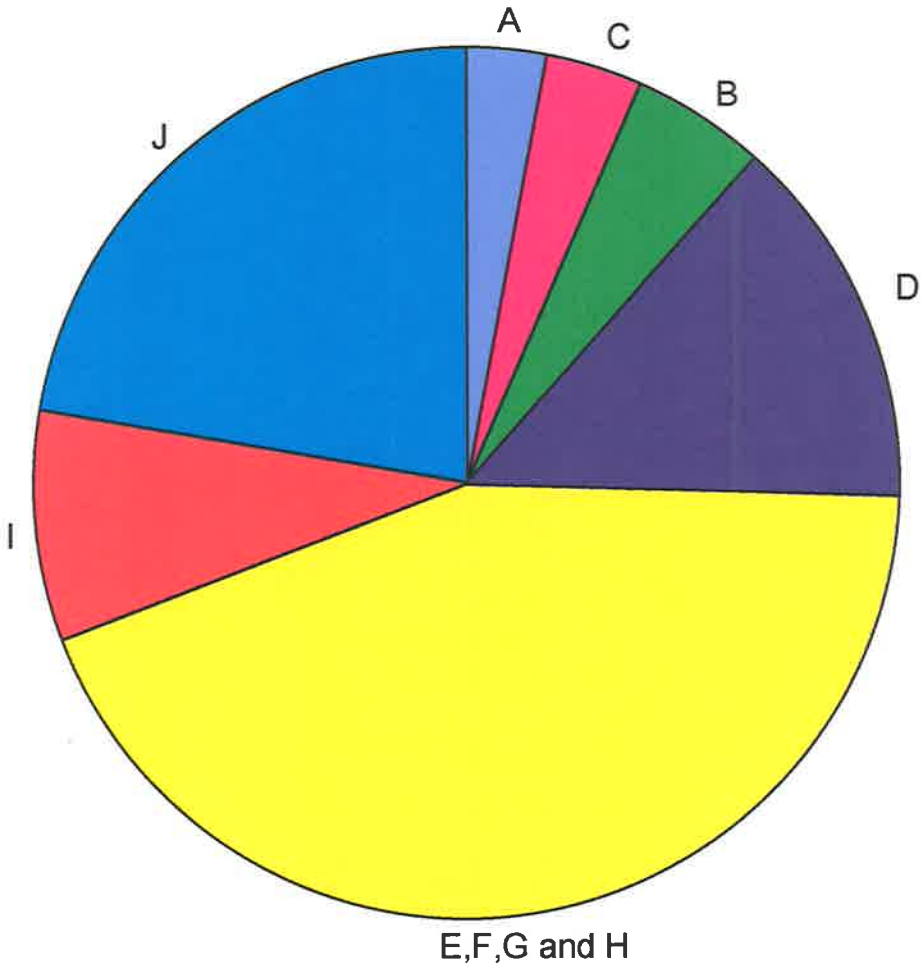
D. Proteins involved in translation

E, F, G, H. House-keeping proteins

I. Signalling and regulatory molecules

J. Others

The percentage of ESTs within each functional category are shown.


A
C
B
J
D
I
E,F,G and H

CATEGORY	No. of ESTs
A. Proteins involved in DNA structure, replication and repair	**16**
i. General	11
ii. Histones/chromatin associated factors	5
B. Transcription factors and associated proteins	**26**
C. RNA-binding proteins and proteins involved in splicing	**18**
D. Proteins involved in translation	**70**
i. ribosomal	60
ii. non-ribosomal	10
E. Cytoskeletal proteins and their regulators	**55**
F. Metabolic proteins	**96**
i. general metabolic enzymes	71
ii. thioredoxin/gluoredoxins	5
iii. ATPases	11
iv. cytochromes and related proteins	9
G. Protein degradation and processing	**30**
i. proteasomes and its subunits	10
ii. proteases and peptidases	10
iii. ubiquitins	3
iv. chaperone and heat shock proteins	7
H. Proteins involved in transport	**40**
i. ion transport and vitamin binding	16
ii. membrane trafficking and vesicle formation	19
ii. nuclear import and export	5
I. Signalling and regulatory molecules	**44**
i. extracellular and membrane-bound	18
ii. intracellular	22
iii. calmodulins	4
J. Others	**114**
I. fluorescent	3
ii. proteins involved in cell-cycle regulation	5
iii. miscellaneous	8
iv. unknown function	38
v. hypothetical	60
Total	**509**

Table 5.1. Summary of EST sequence categories.

In total, 509 ESTs were recognised using a minimum likelihood score (E) of 10^{-6}. The ESTs have been divided, on the basis of function, into 9 categories (left-hand column; A-J). All categories, bar B, C and E have been further sub-divided. The number of ESTs present in each class and sub-class is shown in the right-hand column.

A considerable number of enzymes identified were involved in general metabolism. These included 3 methyl transferases and 5 glutaredoxin/thioredoxin enzymes. 11 sequences matched ATPases known to be involved in general metabolism, and 9 cytochrome and cytochrome-related proteins were isolated. Within the protein degradation and processing category there were 10 sequences which matched known proteasome subunit sequences, 10 proteases and peptidases, 7 heat-shock and chaperone proteins, and 3 ubiquitin proteins, one of which matched the *A. millepora ub52* gene, already submitted to the database. 12 sequences homologous to genes known to play a role in transport of molecules were present, including several Ca^{2+} channel proteins, two amino-acid transporter proteins, a copper transport protein, a potassium channel protein, a sodium/potassium chloride co-transporter, and voltage dependent channels. A potentially interesting find was an EST whose sequence was similar to that of frequenin, a Ca^{2+}-binding protein important in the regulation of neurotransmitter potassium channels in higher animals. Four ion/vitamin-binding proteins were also identified, including ferritin and soma ferritin.

Proteins involved in membrane trafficking were well represented, and included a clathrin homolog, an adaptin homolog, several protein transport sec subunits, translocon-associated proteins, 2 coatomer subunits, and 2 signal recognition particles. Other interesting sequences included a Ca^{2+}-dependent secretion-activator protein and a putative Rab5-interacting protein - a transmembrane protein that may function in endocytic vesicle transport as a receptor for Rab5-GDP. Several proteins were identified that were involved in nuclear import and export, including 2 GTP-binding Ran proteins and a Ran-binding protein.

Several DNA repair proteins, such as a brain my036 protein homolog, a Rad23 protein and a breakpoint cluster region protein, were identified. 8 proteins involved in RNA splicing, along with 9 additional RNA-binding proteins, were also isolated. Further, 3 histone variants, 5 proteins involved in translation initiation, 2 proteins involved in translation elongation, and 2 proteins involved in polyadenylation of mRNA, were represented. 5 proteins were classified as being specifically related to cell cycle regulation, and, interestingly, 3 distinct fluorescent proteins were identified. As corals are not reported to fluoresce the significance of these proteins is unclear.

From a developmental point of view, the most interesting categories were those encompassing the signalling/regulatory molecules and transcription factors. An EGF-motif-containing protein was identified, as was an EST related to the Notch ligand,

Delta. A potential FGF receptor, along with an FGF ligand, a tyrosine kinase receptor and an insulin-like growth factor receptor, were also isolated. In addition, two distinct thrombospondin proteins and an extracellular matrix protein involved in latent TGF-β regulation were represented. A number of intracellular regulatory molecules were identified, including various kinases and phosphatases, in addition to 4 calmodulins, and several Ca^{2+}-dependent proteins. Encompassed in the transcription factor class were both gene-specific and gene non-specific proteins. Several of these, for example an ets-domain protein, hand2′ and a y-box transcription factor, still have unknown functions. However, some of the transcription factors isolated are known to play essential roles during embryonic development, including Enhancer of Zeste, Churchill, Hey and the homeobox proteins, Cnox3 and Hex (see Section 5.4.1 and Chapter 6).

An analysis incorporating all 509 ESTs (see Section 2.2.33.2) illustrates that, in general, the coral ESTs match significantly higher to their chordate homologs than to *D. melanogaster* or *C. elegans* genes (see Section 5.4.1).

5.3 Discussion

This chapter reports the isolation of 3000 *A. millepora* ESTs. After analysis of the 3′ DNA sequence, 509 ESTs were identified as having open reading frames with significant similarity to those of proteins in the database. These predicted proteins were grouped together on the basis of function (see Figure 5.2 and Table 5.1). Due to the equalisation step the percentage of ESTs in each functional category cannot be used to infer their abundance in the corresponding mRNA in *A. millepora*, 96-hours post-fertilisation.

During the isolation of cDNAs for this study, to ensure sequence reads of adequate length, only DNA inserts greater than 500 bp were selected for analysis. However, results indicated that several sequence reads were shorter than 500 bp. This was probably due to inaccurate estimates of fragment sizes from gel electrophoresis. The short lengths of many sequences may mean that sequence similarities correspond to specific protein domains, rather than to particular proteins. Thus, while identified sequences will be referred to here as homologs, in the future it will be important to determine the complete sequence of the ESTs to confirm whether the DNA similarities are truly indicative of homologous genes. Further, it is likely that some ESTs shared no significant similarity to known genes because of short sequence reads that only encompassed 3′ untranslated regions.

It may be expected that as a primitive animal with a single body axis, two body layers composed of relatively few cell types, and a simple nerve net, *A. millepora* would require only a fraction of the genes used to specify the much more complex higher metazoan body plan. If this were true, the proportion of essential protein synthesis and house-keeping genes would be higher than the more complex organisms. However, the array of ESTs isolated in this study (see Appendix A.4), along with the surprising diversity of *A. millepora* genes already identified (reviewed by Ball *et al.*, in press) challenges this. For example, 60 distinct ribosomal proteins are represented in this analysis. This is 11.8% of the total identified ESTs, a value similar to other EST projects (Wan *et al.*, 1996; Ajioka *et al.*, 1998; Suzuki and Satoh, 2000). The number of house-keeping proteins (categories E, F, G and H) totalled 43.6%. This is similar to the 39% identified in the EST analysis of sea urchin embryos performed by Lee *et al.* (1999).

A wide variety of cytoskeletal components were identified in this analysis (see Table 5.1). These include tubulins, kinesins, myosins, actins and a wide range of other cytoskeletal proteins. Not only are all these proteins essential for a diverse range of essential functions involved in maintaining cell structure, but they also play a role in the cytoplasmic transport of organelles and vesicles, and the movement of chromosomes during cell division (reviewed by Huitorel, 1988). In addition, many cytoskeletal molecules function during microtubule-dependent cilia motility (reviewed by Huitorel, 1988). The presence of many microtubule components, then, is consistent with the pre-settlement stage *A. millepora* ectoderm being lined with cilia to allow active swimming of the larvae. As cilia movement functions are regulated by the concentration of Ca^{2+} within the cell (reviewed by Huitorel, 1988), it is not unexpected to see several calmodulin isoforms, and proteins involved in Ca^{2+} transport, represented in this analysis. In addition, the variety of these former-mentioned proteins may also be related to the large amount of Ca^{2+} metabolism that occurs in the adult coral.

Only one identified EST matched an *A. millepora* sequence present in the database. This is not unexpected due to the low number of *A. millepora* cDNAs presently submitted to the databases. Further, it is not surprising that the EST corresponded to a highly abundant *A. millepora* gene, namely *ub52*. This observation highlights the value of the EST approach to gene discovery and characterisation in *A. millepora*.

Several interesting signalling/regulatory proteins and developmentally important transcription factors were identified in this analysis. The presence of these genes in the EST project has implications for their function in the pre-settlement *A. millepora* larva. In higher metazoans, signalling pathways are highly diverse, with signalling superfamilies encompassing a number of protein classes, each of which plays multiple functional roles during development. However, vertebrates use many more signalling molecules than invertebrates. As most or all of these broad classes of signalling molecules are also present in the invertebrates, the vertebrate complexity is likely to result from duplications that occurred in the chordate lineage (Holland, 1998). Phylogenetic analysis of the *A. millepora* signalling molecules identified in this study could provide insight into the complexity of these proteins in what may represent a more ancestral-like animal, and thus help establish how the wide variety of signal transduction cascades present in each of the higher metazoans arose during evolution. Proteins belonging to several classes of signalling molecules were identified here, including FGF, EGF, and TGF-β, along with a tyrosine kinase receptor, a TGF-β receptor, and an insulin growth-like receptor. Genes in the Notch pathway have been well characterised in higher metazoans and function to prevent neural determination (reviewed by Baker, 2000). A homolog of the Notch ligand Delta, was represented in this analysis. Although computer analysis predicted that the probability of this EST being a true homolog was low (E (expected) values = -7), manual examination demonstrated that the alignment looked significant. Further, a homolog of Hey, a transcription factor that acts through the Notch pathway and is important in neurogenesis, organogenesis and somitogenesis (Leimeister *et al.*, 1999; Steidl *et al.*, 2000), was isolated. A homolog of Pecanex, too, was identified (Gilbert *et al.*, 1992). This is a large membrane-spanning protein that has also been implicated in the Notch pathway (LaBonne, 1989a and b).

Coral possesses a primitive nerve net, so analysis of the expression pattern of *A. millepora* neural-specific genes may give some clue to the primitive function of these genes in neural patterning, as well as providing essential marker genes for this aspect of *A. millepora* development. Several genes that can potentially be helpful for these analyses include the homolog of the vertebrate-specific transcription factor, Churchill, which has been implicated in neural patterning in vertebrates (Claudio Stern, personal comm.). Further, Homer-2d, a homolog of which was identified in this study, is enriched at synapses where it binds to group 1 metabotropic glutamate receptors (Kato *et al.*, 1998). Homer proteins have been implicated in the structural changes that occur at synapses during long-lasting neuronal plasticity and

development (Kato *et al.*, 1998). A glia maturation factor-beta homolog was also represented. This regulatory kinase has been demonstrated to stimulate differentiation of normal neurons as well as glial cells (Lim *et al.*, 1989). A homolog of frequenin, a Ca^{2+}-binding protein that modulates the Kv4 channel during neurotransmission (Nakamura *et al.*, 2001), was identified, as was a homolog of the closely related Ca^{2+}-binding protein, neurocalcin. The latter, however, does not have neural function, but interacts with actin in a Ca^{2+}-dependent manner (Mornet and Bonet-Kerrache, 2001). The use of homologs of Pecanex and other members of the Notch pathway (see above), mutants of which show an increase in neural precursor cells, may provide additional insight into the specification of neural fate.

The common ancestor of the protostomes and deuerostomes existed after the separation of the cnidarian lineage. Therefore, it is expected that coral sequences should be equally divergent from sequences from either group. However, in general, coral genes were found to match significantly better with their chordate homologs than with *D. melanogaster* or *C. elegans* genes. This may be explained by the recent, rapid specialisation events that have occurred *D. melanogaster* which have led to a greater divergence of fly genes. An additional explanation stems from the fact that the complexity in vertebrates can be understood in terms of genome wide duplications (Holland, 1998). It is therefore possible that in the chordate lineage ancestral genes have remained relatively conserved, whilst the duplicated genes have diverged to create the complexity of the vertebrates seen today. Conversely, in *D. melanogaster,* where duplication is less prominent, specialisation may have been achieved by the divergence of the original genes from the ancestral state.

It is currently assumed that vertebrate-specific genes, where no protostome homologs have been identified, have recent origins. However, one of the most exciting revelations from this study is that a number of genes previously thought to be vertebrate-specific are in fact present in *A. millepora.* This would suggest that the above assumption is incorrect. The implication of this is clear. With the exception of homologs arising through duplication events, far fewer genes are likely to have been vertebrate innovations than has been assumed to date. Instead, gene loss is likely to have been more extensive in both *D. melanogaster* and *C. elegans* than was previously suspected. Churchill and the methyl CpG-binding protein are two such examples of this phenomenon.

Many of the ESTs have no significant database matches, and a number of ESTs produced matches to proteins of unknown function. To determine whether these ESTs represent untranslated sequence or whether they represent divergent or novel open reading frames it will be necessary to perform a statistical search for open reading frames. When this was performed by Lee *et al.* (1999) they found that between 65-80% of the unidentified ESTs were in fact protein coding sequences. It is possible that a similarly high number of unidentified ESTs reported here will contain coding sequences.

This study has paved the way for a variety of future experiments, the most exciting of which would be an *A. millepora* microarray project. Such a project would analyse gene expression at different stages of *A. millepora* development. Specific questions regarding the switching on and off of genes during, for example, coral bleaching and re-differentiation of the adult coral, could potentially be addressed. In addition, by careful dissection of *A. millepora* embryos during the mass spawning analysis, RNA can be collected from both the oral and aboral ends of the larvae, and used as microarray probes to identify genes expressed along this axis in a spatially restricted manner. This may potentially help answer the questions concerning coral axis specification that were discussed in chapters 3 and 4.

In conclusion, reported here is the initial analysis of 3000 *A. millepora* ESTs. In quantitative terms this is a small EST project, nevertheless it is remarkable that such a number of interesting genes were isolated.

6. *A. millepora hex*: a potential molecular marker of coral development

6.1 Introduction

In the search for potential marker genes a PCR product whose deduced amino acid sequence had high similarity to Hex (Bedford *et al.*, 1993), a homeodomain-containing protein, was identified (see Chapter 5). Homeodomains are 60 amino acid DNA-binding motifs found in a wide variety of transcription factors that play fundamental roles in cell differentiation during development (Gehring, 1987). A number of genes that encode homeodomain-containing proteins have already been identified in the coral. These include several *pax* genes, *eve, emx1*, *cnox1* and *cnox2* (reviewed by Ball *et al.*, in press). In addition to these genes, one of the ESTs (see Appeddix A.4) showed significant similarity to *cnox3* (Schummer, 1992). Based on sequence relatedness, homeobox-containing genes can be divided into two classes, the Antennapedia (Antp) class and the Paired class. The *hox*, *parahox* and *non-hox* gene families constitute the Antp class, the *non-hox* genes being the most divergent group (Gauchat *et al.*, 2000). Hex belongs to the Antp class of non-hox homeodomain proteins.

To date, *hex* genes have been identified in both vertebrates (for example, Thomas *et al.*, 1998; Jones *et al.*, 1999; Ho *et al.*, 1999) and *Hydra* (Gauchat *et al.*, 2000), but no protostome homolog is known. In the mouse, Hex is the earliest molecular marker for anterior development (Thomas *et al.*, 1998; see Section 6.4). In addition, *hex* is expressed in the definitive endoderm of the mouse and, later in embryogenesis, activated in some endoderm-derived tissues, as well as endothelial and hematopoietic precursor cells (Thomas *et al.*, 1998; see Section 6.4). Similar findings have been documented in other vertebrates (for example, Jones *et al.*, 1999; Ho *et al.*, 2000; Brickman *et al.*, 2000). The role Hex plays during early anterior signalling, its early tissue-specific expression in the endoderm, and its later role in differentiation of mesoderm-derived tissues, prompted a more thorough analysis of this gene as a potential *A. millepora* tissue or axis-specific molecular marker.

6.2 *D. melanogaster* CG7056, a putative *D. melanogaster hex* gene

6.2.1 Identification of *D. melanogaster hex*

Previously it has been documented that *hex* is a vertebrate-specific gene, with no protostome homolog (see Section 6.1). However, it was important to confirm this because the presence of *hex* homologs expressed in a tissue-specific fashion in the protostome lineage would increase the likelihood that *hex-Am* could provide a candidate molecular marker.

The *X. laevis* Hex sequence was therefore used as a query to search sequence databases (see Section 2.2.33.1) for invertebrate gene products showing sequence similarity. The *D. melanogaster* sequence with the most significant similarity to *X. laevis* Hex was the gene product CG7056 (Genbank #AE003733, see Appendix A.6). The deduced amino acid sequence of CG7056 was 23.9% identical to *X. laevis* Hex (Figure 6.1A). A *C. elegans* gene product, M6.3, of similar sequence similarity (23.7% identical to *X. laevis* Hex) was also identified (see Figure 6.1A). Although these values of similarity are low, identities with *X. laevis* Hex were substantially higher within the homeodomain, at 69.2% and 61.5% for *D. melanogaster* CG7056 and *C. elegans* M6.3, respectively (see Figure 6.1B). Further, phylogenetic analyses of the homeodomain sequences of *D. melanogaster* CG7056, *C. elegans* M6.3 and other Hex proteins, with homeodomain sequences from a variety of homeodomain containing transcription factors, clearly identifies both *D. melanogaster* CG7056 and *C. elegans* M6.3 as Hex proteins (see Figure 6.7). Together, these results demonstrate that the uncharacterised *D. melanogaster* gene product CG7056 is a diverged *hex* gene. It was therefore decided to analyse the expression pattern of this gene to see if it showed any similarity to that of vertebrate *hex*.

6.2.2 Spatial expression pattern of *D. melanogaster hex* during embryogenesis

Comparison of the CG7056 sequence with *D. melanogaster* genomic DNA sequences indicated that the 819 bp coding region of *hex* was organised as two exons, separated by a 3.861 kb intron (see Appendix A.6). To analyse the spatial expression pattern of *hex*, *in situ* hybridisation was performed. Initially a construct was created which contained sequences from the *hex* coding region 3′ to the homeobox, plus 3′ untranslated sequences. Briefly, the sequences were PCR amplified from genomic DNA (see Section 2.2.8) using specific primers (Hexinsitu1 and 2; see Sections 2.1.12.9 and 2.2.10.1). This generated a 579 bp fragment, which was inserted into the pGEM-T® Easy vector (see Sections 2.2.5-2.2.7). Sequence

A.

Hex-*Am*	*Xenopus* Hex	Rat Hex	Mouse Hex	Human Hex	*Hydra* Hex	CG7056 *Dm*	M6.3 *Ce*	
	37.8	33.7	33.7	36.3	28.9	28	28	Hex-*Am*
		65.7	66.8	67.4	29.5	23.9	23.7	*Xenopus* Hex
			97	93	28.9	24	23.3	Rat Hex
				91.5	28.9	24.4	23.3	Mouse Hex
					28.9	23.7	23.3	Human Hex
						23.1	26.6	*Hydra* Hex
							22.6	CG7056-*Dm*
								M6.3 *Ce*

B.

Hex-*Am*	*Xenopus* Hex	Rat Hex	Mouse Hex	Human Hex	*Hydra* Hex	CG7056 *Dm*	M6.3 *Ce*	
	81.7	81.7	81.7	81.7	66.7	65	63.3	Hex-*Am*
		100	98.3	100	65	66.7	66.7	*Xenopus* Hex
			98.3	100	65	66.7	66.7	Rat Hex
				98.3	65	66.7	66.7	Mouse Hex
					65	66.7	66.7	Human Hex
						51.7	55	*Hydra* Hex
							60	CG7056-*Dm*
								M6.3 *Ce*

Figure 6.1. Sequence identities between various Hex proteins from different organisms.

Amino acid sequences were aligned using CLUSTALW and the identities between the sequences are shown as percentage values. A: identities for the entire length of the polypeptide. B: identities for the homeodomain region only. Note that the extreme 5' region of the *Hydra* Hex polypeptide sequence is yet to be determined. *Dm* = *D. melanogaster*; *Ce* = *C. elegans*.

analysis confirmed the orientation of the fragment, which was designated pGEM-T-*hexDm3′*. pGEM T-*hexDm3′* was linearised with the restriction enzyme *Pst*I, and subsequently used as a template to generate a DIG-labelled RNA probe for whole mount *in situ* hybridisation on *D. melanogaster* embryos (see Section 2.2.25). Although a specific expression pattern could be detected, the data were ambiguous due to the presence of non-specific staining. Thus, a second construct was designed. Briefly, the entire coding region of *hex*, plus the 3.861 kb intron, was PCR amplified using specific primers (HexORF1 and HexORF2; see Section 2.1.12.9). This yielded a single product of 4.68 kb, which was inserted into the pGEM-T® Easy vector (see Sections 2.2.5-2.2.7). The construct was designated pGEM-T-*hexDmgen*. Restriction digestion of this construct with *Hpa*II cut the DNA at two sites, 74 bp and 19 bp into the 5′ and 3′ end of the intron, respectively. This excised a DNA fragment of 3.768 kb, corresponding to almost the complete intron sequence. The remainder of the construct (pGEM-T® Easy plus 912 bp *hex* coding region insert) was re-ligated, and subsequently designated pGEM-T-*hexDm*. pGEM-T-*hexDm* was linearised with the restriction enzyme *Sal*I, and used as a template to generate a DIG-labelled RNA probe for whole mount *in situ* hybridisation on *D. melanogaster* embryos (see Section 2.2.25). The probe showed an identical staining pattern to that observed with the first probe but without the non-specific staining (Figure 6.2).

hex transcripts are initially detected in the syncitial blastoderm (Figure 6.2A), suggesting a maternal deposition of *hex* mRNA, after which transcripts were not detected until stage 11 of *D. melanogaster* embryogenesis. At this stage gastrulation is complete, the germband is fully extended and *hex* transcripts can be specifically identified in the region of the posterior midgut primordium (Figure 6.2B). As the embryo develops and the germband shortens, *hex* expression is maintained in the region of the posterior midgut primordium (Figure 6.2C). At stage 13 the anterior and posterior midgut primordia fuse. *hex* expression was seen, and indeed appeared to be intensified, at the point of fusion (Figure 6.2D). After fusion, the midgut closes to encompass the yolk sac, and then constricts three separate times along the anterior/posterior axis (Bienz, 1994). The second midgut constriction occurs, initially, at approximately stage 15, and *hex* transcripts can be identified at and surrounding this constriction site (Figure 6.2E). *hex* expression persists during the remainder of embryogenesis, being specifically located between the first and third constriction sites (Figure 6.2F). In addition to the expression of *hex* in the endoderm, at stage 11, prolonged development of the *in situ* hybridisation colour detection assay revealed an apparently lower activation of *hex* in the mesoderm. As this was only detected in

Figure 6.2. *hex* expression during *D. melanogaster* embryogenesis.

Whole mount *in situ* hybridisation was performed on wild-type *D. melanogaster* embryos, at various stages of development, using RNA DIG-labelled *D. melanogaster hex* as a probe.

(A) **Syncitial embryo**: Maternally deposited *hex* transcripts are evident.

(B) **Stage 11**: *hex* transcripts are detected specifically in the region of the posterior midgut primordium.

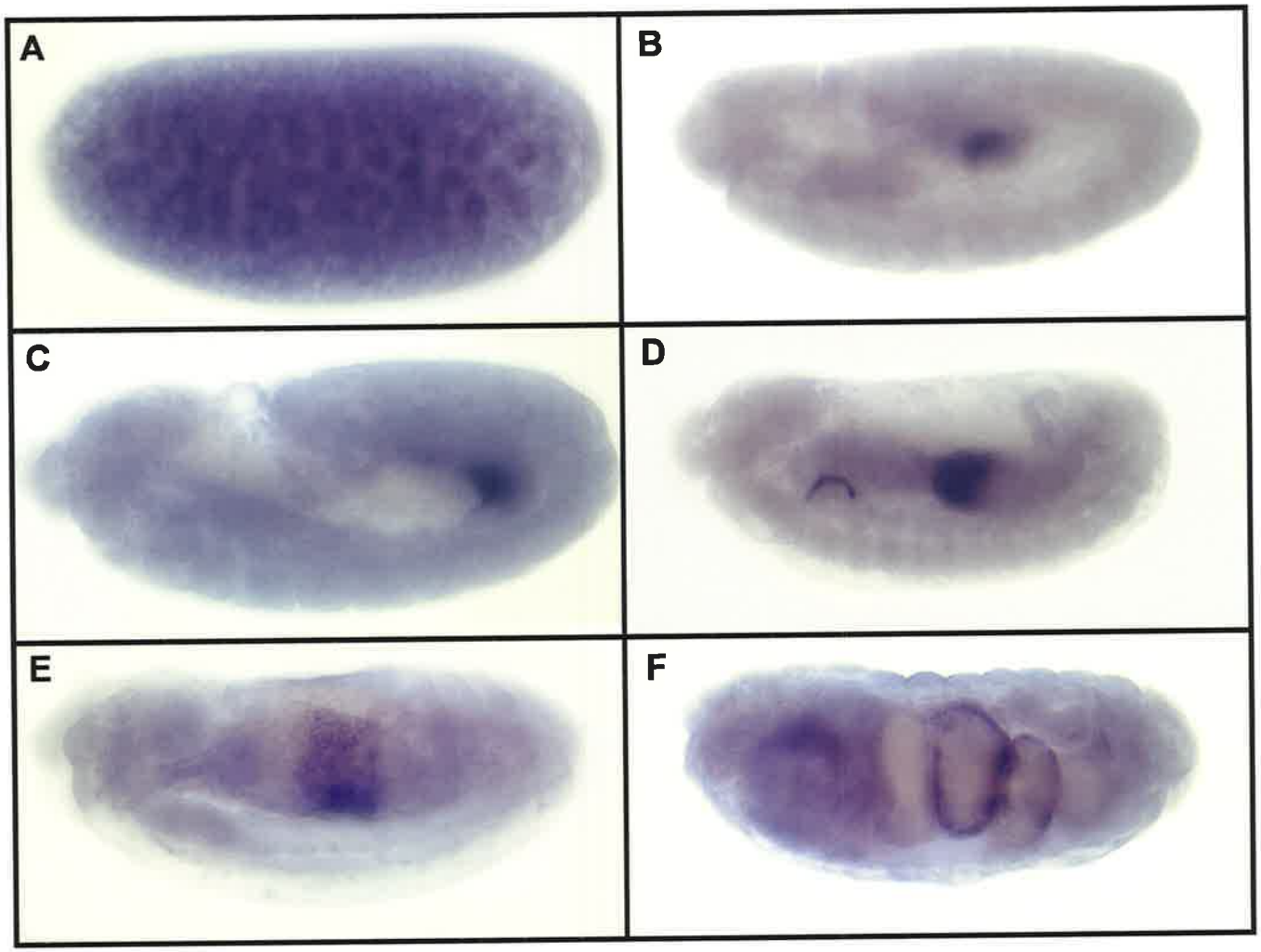
(C) **Stage 12**: *hex* transcripts remain confined to cells in the region of the posterior midgut primordium.

(D) **Stage 13**: *hex* expression is intensified at the point of fusion of the anterior and posterior midgut primordia.

(E) **Stage 15**: *hex* transcripts can be identified at, and surrounding, the site of the second midgut constriction.

(F) **Stage 16**: *hex* transcripts are specifically located between the first and third constriction site in the fully formed midgut.

Anterior left, lateral view.


A
B
C
D
E
F

approximately half the embryos analysed it was unclear if this result was true or represented background staining (data not shown).

6.2.3 A fly stock with a *hex* deficiency

When studying previously uncharacterised genes, it is advantageous to have several controls when analysing expression data, especially when the expression pattern is ambiguous (see above: *hex* expression in the mesoderm). Ideally this would be a null mutant for the gene of interest. For *D. melanogaster hex* there is no null mutant available. However fly stock 3340 (*Df(3R)eR1, Ki^{1}/ TM3, Sb^{1}Ser^{1}*) has a homozygous lethal deficiency in the third chromosome near, and possibly incorporating, the chromosomal position of CG7056. Flies in this stock are maintained over a dominantly marked balancer chromosome (see Section 2.1.9). It was predicted that if the *hex* sequences were absent in flies heterozygous for this deficiency, then these flies would have half the amount of genomic *hex* DNA to wild-type flies. Thus, to determine whether the *hex* sequence was deleted, a genomic southern analysis was performed. Specifically, genomic DNA was extracted from adult flies (see Section 2.2.8) and digested with *Eco*RI. Two probes were generated for southern analysis. Briefly, pGEM-T-*Dmhex3′* (see Section 6.2.2) was digested with *Eco*RI, and the 579 bp DNA fragment corresponding to the *Dmhex3′* sequence was isolated and used to generate the first probe. The second probe was generated from a 300 bp DNA fragment corresponding to *D. melanogaster cyclinE* (courtesy of Donna Crack). The *D. melanogaster cyclinE* gene is located on the second chromosome, and since it should be present equally in both the deficiency heterozygotes and the wild-type flies, this probe acted as a loading control. The result of the southern hybridisation is shown in Figure 6.3. As expected, the *cyclinE* DNA probe hybridised to a 3.3 kb genomic fragment, and the *hex* probe hybridised to a genomic fragment of approximately 6 kb. The intensity of each of the radioactive signals was determined using a phosphorimager (see Section 2.2.15) and, accounting for the intensities of the *cyclinE* hybridising bands, the ratio of *hex* signal intensity between the deficiency flies and wild-type flies was 1:1.87. This figure confirmed that the *hex* sequence was absent in flies containing the third chromosome deficiency, and thus provided a suitable negative control for the *hex* expression pattern seen in *D. melanogaster* embryos.

To determine whether the *hex* expression in the mesoderm was true, *in situ* hybridisation was performed using deficiency stock embryos. Although the above results predicted that the *hex* probe would only hybridise to three quarters of the

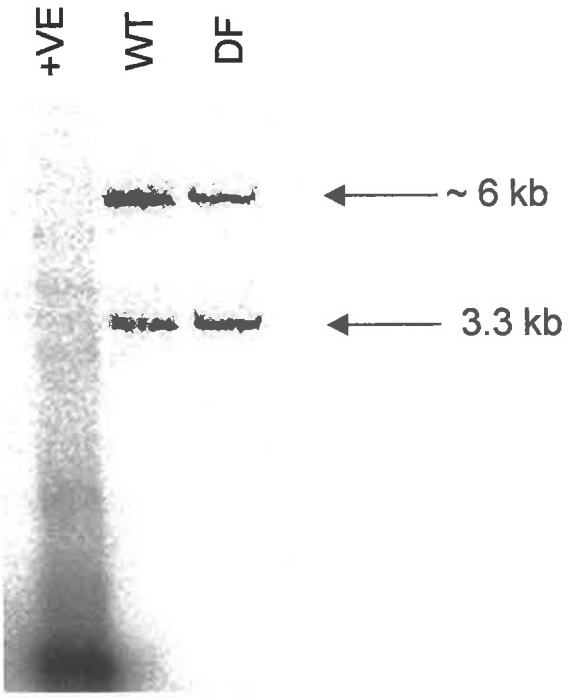


Figure 6.3. Southern hybridisation of genomic DNA from both wild-type flies and deficiency stock flies.

DNA was loaded onto a 1% agarose gel as follows:

Lane 1: *hex* cDNA positive control (+VE)

Lane 2: *Eco*RI-digested genomic DNA from wild-type flies (WT)

Lane 3: *Eco*RI-digested genomic DNA from deficiency stock flies (DF)

The DNA was subjected to electrophoresis and blotted to nylon filters. The filters were probed with the appropriate ^{32}P-labelled DNA (see text). Sizes of resulting bands were estimated based on migration of marker DNA on the same gel. In both the wild-type and the deficiency stock flies, a 3.3 kb band can be seen hybridising to the *cyclinE* DNA probe, and an approximately 6 kb band can be seen hybridising to the *hex* DNA probe. Analysis of the intensities of each band demonstrated that *D. melanogaster hex* is absent in the deficiency stock flies.

Figure 6.4. Expression of *labial* during midgut development in *D. melanogaster* embryos.

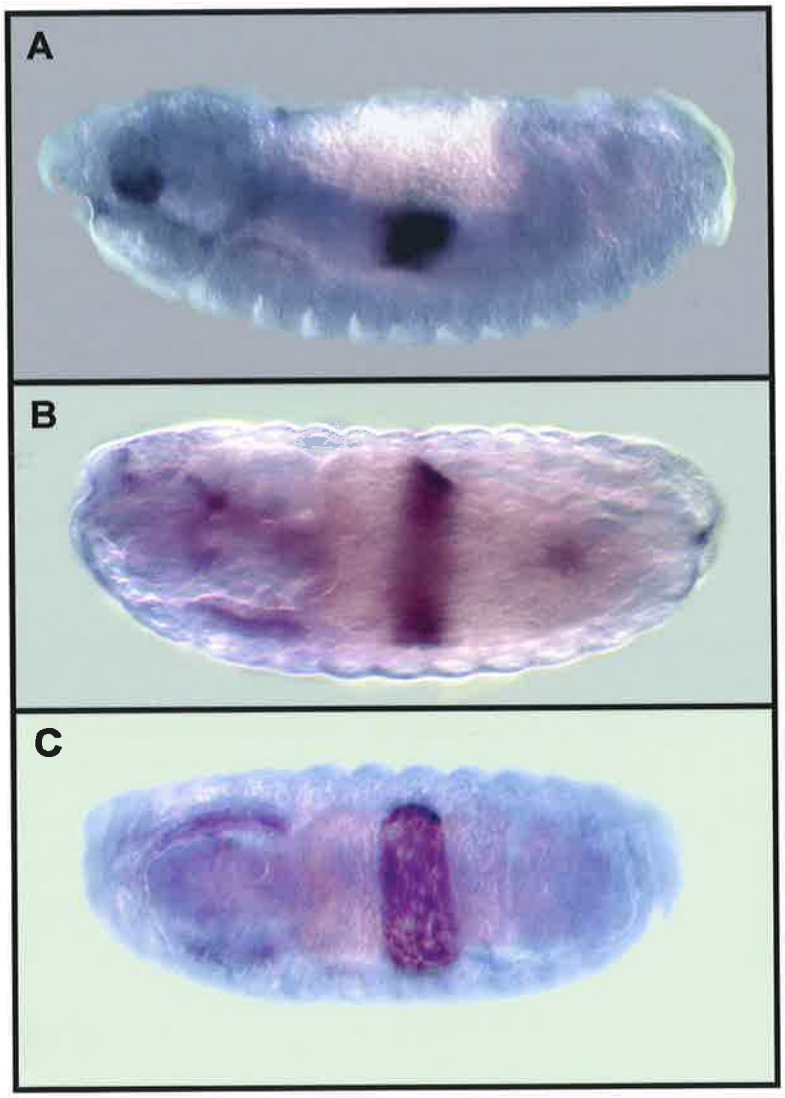
Whole mount *in situ* hybridisation was performed on wild-type *D. melanogaster* embryos, at various stages of development, using *D. melanogaster labial* DIG-labelled RNA as a probe.

(A) **Stage 13**: Similar to *hex*, *labial* is expressed at the point of fusion of the anterior and posterior midgut primordia. Lateral view.

(B) **Stage 15**: *labial* transcripts can be identified at the site of the second midgut constriction, in a narrower band than the *hex* transcripts. Dorsal/Lateral view.

(C) **Stage 16**: Unlike *hex*, *labial* transcripts are only located between the first and second constriction site in the fully formed midgut. Dorsal up.

Anterior left.


A
B
C

embryos, the quality of the expression data were insufficient to allow a statistical analysis. Thus, to enable the embryos heterozygous for *hex* to be distinguished during the *in situ* hybridisation experiment, it was necessary to place the deficiency stock of flies over a marked balancer chromosome. Specifically, the balancer chromosome incorporated an *Ultrabithorax* (*Ubx*)*lacZ P*-element (see Sections 2.1.9 and 3.3.1). This allows transcription of the *lacZ* gene to be driven by the *Ubx* promoter in heterozygous embryos, such that *lacZ* would exhibit an expression pattern identical to a proportion of the endogenous *Ubx* expression pattern, allowing identification of the homozygous deficiency embryos by the absence of this expression pattern. Unfortunately, although attempted several times, immunohistochemical analysis (see Section 2.2.26) using an anti-β Galactosidase antibody (see Section 2.1.7.1) showed no specific expression pattern in any embryos examined (see Section 6.4). Because of time constraints for this thesis, no further work aimed at confirming the mesodermal expression of *hex* was performed.

The expression pattern of *D. melanogaster hex* indicates that, unlike vertebrate *hex*, this gene does not seem to act as an anterior marker during early development. Although it is still unknown whether the *hex* expression seen in the *D. melanogaster* mesoderm is true, *hex* endoderm-specific expression is seen in both *D. melanogaster* and vertebrates raising the possibility that *hex* expression could act as an *A. millepora* endodermal marker gene.

6.3 Characterisation of *A. millepora hex, hex-Am*

6.3.1 Isolation of *hex-Am*

DNA prepared (see Section 2.2.7.2) from the bacterial stock incorporating the *hex-Am* EST (position E3, plate 1; see Section 5.2.1) was digested with the restriction enzymes *Eco*RI and *Xho*I (see Section 2.2.1/3/4). This yielded a 250 bp DNA fragment, which corresponded to the *hex-Am* PCR product. This DNA fragment was subsequently used as a probe to screen 50,000 plaques of the *A. millepora* pre-settlement stage cDNA library. The resulting five positive plaques were excised from their λZAP vector and sequence analysis (see Section 2.2.18), using the T7 primer (see Section 2.1.12.1), confirmed that they all originated from the same mRNA species. Subsequently, the complete DNA sequence of the clone with the longest cDNA insert was analysed via a stepwise approach, in that new primers were designed from each sequencing result (see Section 2.1.12.18). The 3,570 bp clone, referred to as *hex-Am*, contains a complete open reading frame plus 5′ and 3′ untranslated sequences (Figure 6.5). It has a polyadenylation signal (AATAAA) 16 bp upstream of the polyA tail

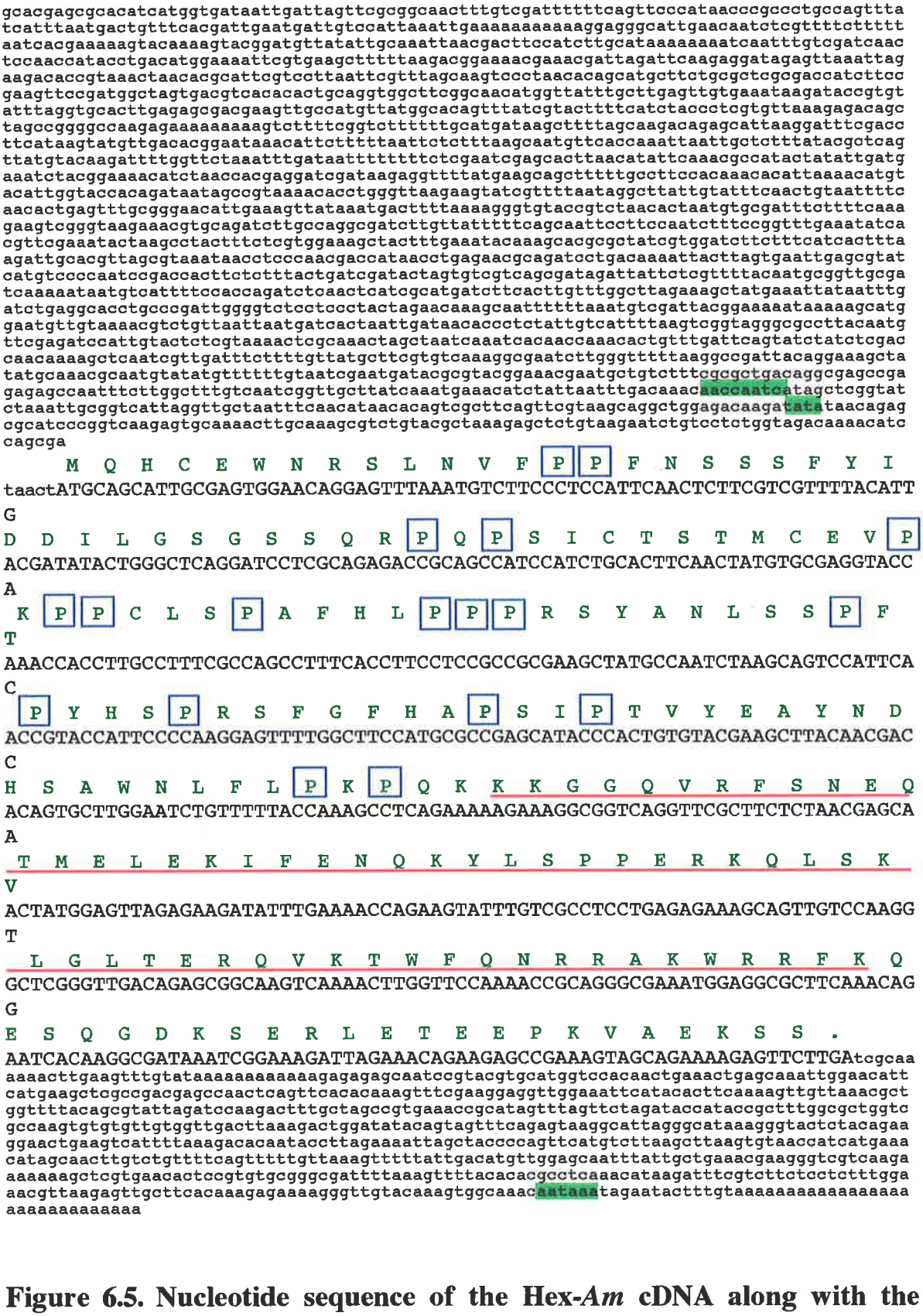


Figure 6.5. Nucleotide sequence of the Hex-*Am* cDNA along with the deduced amino acid sequence of the coding region.

Non-coding nucleotides are represented with lower case letters. Coding nucleotides are represented with capital letters with their corresponding amino acid residues shown in green. The homeobox is indicated by a red underline, the proline residues in the N-terminal domain are boxed in blue. The CAAT box, TATA box and polyadenylation site are shaded in green.

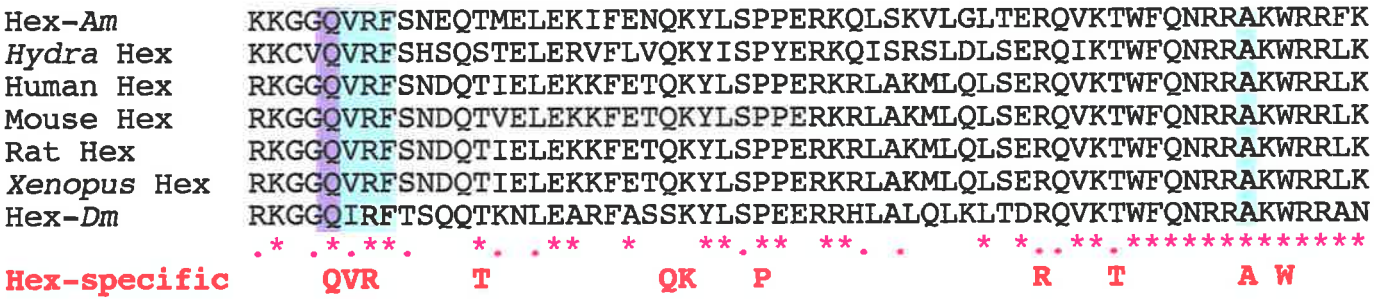


Figure 6.6. Sequence alignment of the Hex-*Am* homeobox with the homeoboxes of its homologs in other organisms. Sequences were aligned using CLUSTALW. *= identical residues; .= conserved residues. Q5 is conserved in all homeoboxes and is shaded in purple. V6, R7, F8 and A54 have been implicated in providing Hex-specific DNA binding and are shaded in blue. Note the conservative substitution of V6 to I6 in Hex-*Dm*. Amino acid residues shown in red are the Hex family-specific residues proposed by Gauchat *et al.* *Dm* = *D. melanogaster*; *Ce* = *C. elegans*.

(Figure 6.5). Predicted translation start and stop sites are positioned at bases 2,274 and 2,855, respectively, giving the predicted protein a length of 194 amino acids and a predicted molecular mass of 22.3 kDa. The homeodomain is positioned between amino acid residues 110 and 170 (see Figure 6.5).

As implied by their original name, (Proline (Pro)-rich homeodomain protein; Crompton *et al.*, 1992), Hex proteins contain a Pro-rich N-terminal domain. In Hex-*Am*, this region is positioned between amino acid residues 1 and 109, and is 17% Pro-rich. Within the N-terminal domain of both Hex-*Am* and its vertebrate homologs, but not *D. melanogaster* Hex, is a TN motif (TPFYIDDIL; see Figure 6.5). This was first discovered in the Nk2 family of transcription factors and is suggested to have a repressor function (Harvey, 1996). However, even though Hex has been demonstrated to execute its roles during development by acting as a transcriptional repressor (Ho *et al.*, 1999; Pellizzari *et al.*, 2000), this TN motif does not seem to be important for this function (Tanaka *et al.*, 1999) and thus its absence in *D. melanogaster* Hex may not be significant.

6.3.2 Comparative analysis of Hex-*Am*

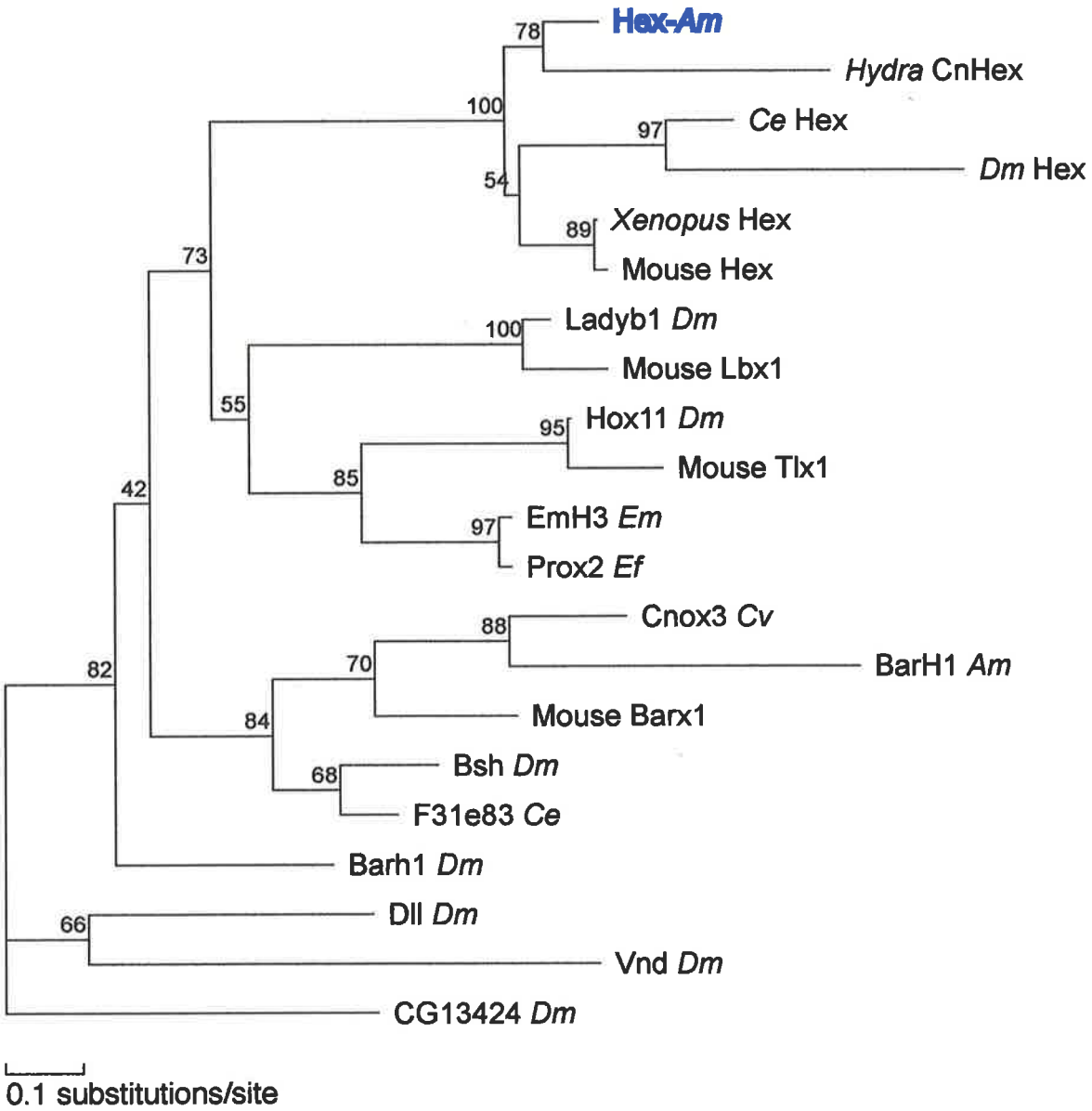
It is interesting to note that the Hex-*Am* polypeptide (194 amino acid residues) is substantially shorter than Hex sequences found in higher metazoans, which are typically around 270 amino acid residues. Further, although the entire coding region of *Hydra hex* is still unknown, based on sequence comparisons it is likely that it will have a sequence length equivalent to that of Hex-*Am*. Within the homeodomain, Hex-*Am* is most similar to vertebrate Hex (83.1% identity with *X. laevis*, human, mouse and rat; see Figures 6.1B and 6.6) while identity to the homeodomain of *D. melanogaster* Hex and *Hydra* Hex is somewhat lower (67.7% and 69.2% identity, respectively; see Figures 6.1B and 6.6).

Ordinarily, within the homeodomain motif there is a conserved arginine (Arg) positioned at residue 5, which has been demonstrated to play a pivotal role during DNA-binding (Gehring *et al.*, 1994). In Hex proteins, and in Hex-*Am*, this Arg has been replaced with a glutamine (Gln), although the significance of this is, as yet, unknown. Residues at positions 6, 7, 8 and 54 of the homeodomain have been implicated in providing family-specific DNA-binding of this motif. In accordance with this, these residues have been conserved between Hex-*Am* and its higher metazoan homologs (Val6, Arg7, Phe8 and Ala54), bar a conservative substitution of Val6 to Ile6 in *D. melanogaster* Hex (see Figure 6.6).

Figure 6.7. Phylogenetic relationships of Hex.

The maximum likelihood analysis was conducted on the amino acid alignment of the homeobox only (Figure 6.6 illustrates the borders of the domain). Numbers in parentheses are GenPep identification numbers. Numbers above branches indicate the percentage of 2000 bootstrap replicates supporting the topology shown.

Dm = D. melanogaster; *Ce = C. elegans*; *Ef = Ephydatia fluviatilis*; *Cv = Chlorohydra viridissima*


Hex-Am
Hydra CnHex
Ce Hex
Dm Hex
Xenopus Hex
Mouse Hex
Ladyb1 Dm
Mouse Lbx1
Hox11 Dm
Mouse Tlx1
EmH3 Em
Prox2 Ef
Cnox3 Cv
BarH1 Am
Mouse Barx1
Bsh Dm
F31e83 Ce
Barh1 Dm
Dll Dm
Vnd Dm
CG13424 Dm
78
100
97
54
89
73
100
55
95
85
97
42
88
70
82
84
68
66
0.1 substitutions/site

Phylogenetic analysis (see Figure 6.7) of the Hex-*Am* homeodomain and various other homeodomain sequences places Hex-*Am* basal to its higher metazoan homologs. As predicted, *Hydra* Hex shows more divergence than Hex-*Am* (see Figure 6.7).

6.3.3 Temporal expression of *hex-Am*

In order to examine the temporal expression of *hex-Am*, a northern hybridisation was performed using mRNA isolated from various stages of *A. millepora* development (see Section 2.2.14). Specifically, the Hex-*Am* cDNA was digested with the *Hin*dIII restriction enzyme. This yielded several DNA fragments. The 731 bp fragment, which corresponded to 303 bp of *hex* coding sequence plus 428 bp of 3′UTR, was isolated and subsequently used to generate a probe (see Section 2.2.12). Figure 6.8 shows a single species hybridising to the partial Hex-*Am* cDNA. *hex* transcripts were not seen during early *A. millepora* development, being initially detected, albeit at a low level, in 48-hour-old embryos. Expression increased in embryos at the pre-settlement stage of development. Although the size of the hybridising band was approximately 1.4 kb, the longest Hex-*Am* cDNA isolated was 3.57 kb (see Section 6.3.1). This Hex-*Am* cDNA may represent a rare additional transcript that was not present on the northern hybridisation. Alternatively, the Hex-*Am* cDNA may correspond to a chimeric clone incorporating both the *hex* sequence, and additional DNA sequences. Consistent with this, there is a putative CAAT box (AACCAATCA) positioned at 2,068 of *hex-Am*, followed by a putative TATA box 95 bp downstream. This predicts a 1.5 kb *hex-Am* transcript, a size similar to that detected during northern analysis.

6.3.4 Spatial expression of *hex-Am*

To analyse the spatial expression pattern of *hex-Am*, the Hex-*Am* cDNA was linearised with the *Acc*I restriction enzyme and used as a template to generate a DIG-labelled RNA probe for *in situ* hybridisation on *A. millepora* embryos (see Sections 2.2.25.1 and 2.2.31). However due to reasons discussed in Section 4.4, this technique was unsuccessful and the spatial expression pattern of *hex-Am* remains to be elucidated.

6.4 Discussion

This chapter reports the analysis of *hex-Am*, and its potential to act as a molecular marker during *A. millepora* development.

When an EST showing similarity to the Hex polypeptide was first identified, the intriguing role of Hex as the earliest anterior marker of vertebrate development, in addition to its early expression in the mouse endoderm, suggested that one or both of these functions may have been conserved in the coral. Here a more in depth

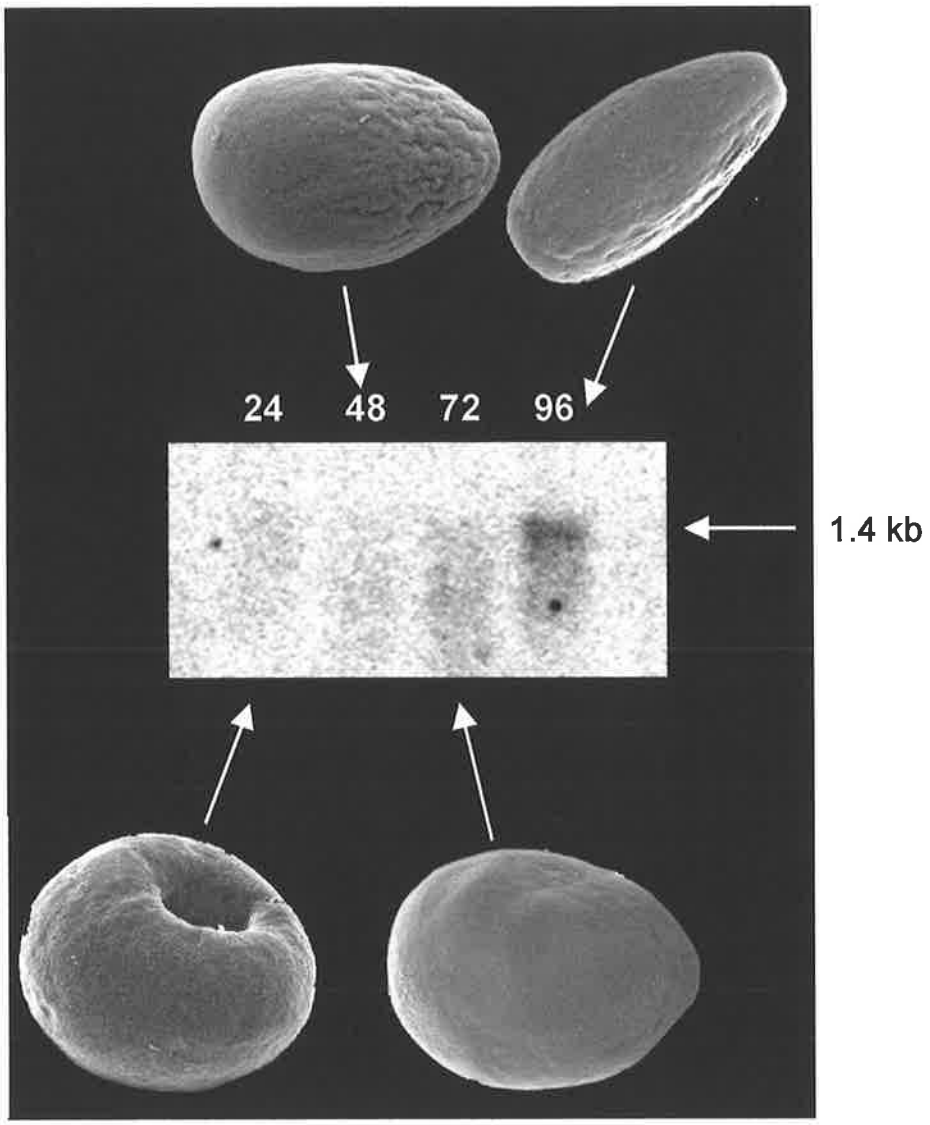


Figure 6.8. Northern hybridisation of *hex-Am* cDNA to *A. millepora* embryonic mRNA.

1μg of embryonic mRNA isolated from different stages of *A. millepora* development (courtesy of Dave Hayward) was subjected to electrophoresis on a 1% agarose gel, and blotted onto a nylon filter. The filter was probed with ^{32}P-labelled *hex-Am.* Sizes of resulting bands were estimated based on migration of marker DNA on the same gel. A single species of approximately 1.4 kb (arrow) can be detected at 96 hours post-fertilisation and very faintly at 72 hours post-fertilisation. Each lane is labelled with a number that represents the age of the embryo, in hours, from which the mRNA was isolated. Electron micrograph pictures (courtesy of Eldon Ball) corresponding to each stage of development are shown.

description of Hex function during mouse embryogenesis is given before returning to the *A. millepora hex* gene.

Gastrulation commences in the mouse at 6.5 days postcoitum (dpc) and is identified by the formation of the primitive streak, a transient embryonic structure that is localised to one side of the epiblast (see Figure 6.9). The primitive streak defines the future posterior of the embryo and, as it is the first morphological evidence of anterior/posterior patterning, it was also believed to be the first marker of anterior/posterior axis specification. However, recent analysis of *hex* has demonstrated that it is the visceral endoderm that first acquires anterior/posterior polarity, subsequently cuing the initial formation of anterior structures (Thomas and Beddington 1996; Thomas *et al.*, 1997; Varlet *et al.*, 1997; Pöpperl *et al.*, 1997). Initially *hex* is expressed in the primitive endoderm of the implanting blastocyst, but at 4.5 dpc *hex* transcripts are confined to a small patch of visceral endoderm cells at the distal most tip of the egg (Thomas *et al.*, 1998; see Figure 6.9). These *hex*-expressing cells are destined to give rise to the anterior visceral endoderm (AVE). They subsequently undergo directional migration and, by 6.5 dpc, have assumed a more anterior position (Thomas *et al.*, 1998; see Figure 6.9). *hex* expression, here in the AVE, marks the earliest molecular anterior/posterior asymmetry in the mouse embryo, over 24 hours before the onset of gastrulation.

As gastrulation proceeds, the primitive streak elongates to the distal end of the embryo. Cells at the anterior tip of the streak represent the progenitors of the node, and will eventually generate mesoderm and definitive gut endoderm. At stage 7, *hex* is activated in the anterior definitive endoderm (ADE) (Tam and Steiner, 1999), which subsequently extends from the node to merge with the AVE (see Figure 6.9). The co-ordination of signals from the AVE and the ADE has been implicated in full anterior neural patterning (Bedington and Robinson, 1999; Shawlot, 1999; reviewed by Martinez Barbera, 2001). Comparative expression analysis suggests that other vertebrates have structures that are functionally analogous to the mouse AVE, all of which express *hex* homologs and regulate the formation of anterior structures (Jones *et al.*, 1999; Ho *et al.*, 1999; Brickman *et al.*, 2000).

Later in mouse embryogenesis, *hex* is expressed in a variety of endodermally-derived tissues (Bogue *et al.*, 2000; Keng *et al.*, 2000; Martinez Barbera *et al.*, 2000; Pellizzari *et al.*, 2000). *hex* is also activated in mouse mesoderm-derived tissues, including nascent blood islands, endothelial cell precursors and multipotent

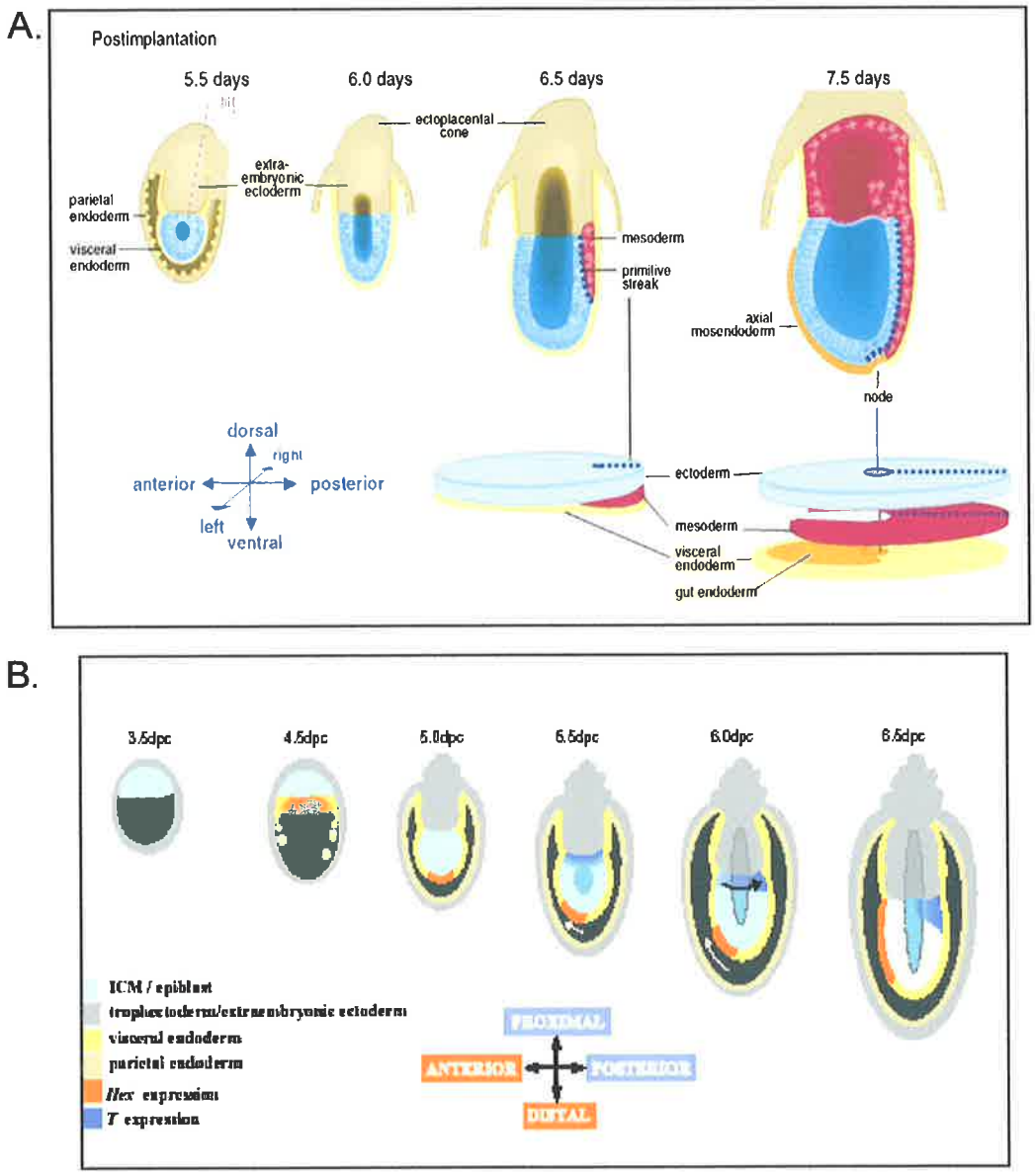


Figure 6.9. Schematic of mouse post-implantation development.

A. The germ layers of the 6.5 and 7.5 day embryos are shown as flattened sheets for ease of understanding and comparison. Blue = ectoderm, red = mesoderm, yellow = visceral endoderm and orange = gut endoderm. The major axes of the embryo are shown superimposed on the profile of an adult mouse. Not drawn to scale. Taken from Beddington and Robertson, 1999.

B. A model for anterior/posterior axis formation in the mouse embryo. *Hex* expression (orange) is initiated in the nascent primitive endoderm of the 4.5 dpc blastocyst and at 5.0 dpc is restricted to a few visceral endoderm cells at the distal tip of the egg cylinder. At 5.5 dpc this proximodistal asymmetry is converted into anteroposterior asymmetry by the movement of the distal *hex*-positive cells (white arrow). Taken from Thomas *et al.*, 1998.

hematopoietic cells (Crompton *et al.*, 1992; Keng *et al.*, 1998; Thomas *et al.*, 1998). *hex* expression in both the endothelial and hematopoietic precursors is distinguished during their terminal differentiation, suggesting a role for this gene during selection and/or initial differentiation steps (Thomas *et al.*, 1998).

hex has previously been reported as a vertebrate-specific gene (Gauchet *et al.*, 2000). However, the homeodomain region of the *D. melanogaster* gene product CG7056 showed 69.2% identity with the *X. laevis* Hex homeodomain (see Figure 6.1B and 6.6). Although this value was much lower than the 83.7% identity between this region of Hex-*Am* and its vertebrate homologs, it was similar to the identity of the previously reported *Hydra* Hex homeodomain (Gauchet *et al.*, 2000) with the homeodomain of both Hex-*Am* (69.2%) and vertebrate Hex (67.2%; see Figure 6.1B and 6.6). Consistent with this, phylogenetic analysis supports the proposal that the *D. melanogaster* gene product CG7056 corresponds to a highly diverged Hex (see Figure 6.7). From similar phylogenetic analyses of *Hydra* and vertebrate Hex, Gauchet *et al.* suggested a Hex-specific signature, based on certain residues within the homeodomain motif that are Hex-specific (see Figure 6.6). Although all these residues are conserved in Hex-*Am*, they are not in *D. melanogaster* Hex. It is important to now incorporate *D. melanogaster* Hex into the family, and possibly refine this signature. It is interesting to note that *hex* is yet another example of an *A. millepora* gene showing more similarity to its vertebrate homologs than to its invertebrate ones (see Section 5.3.1). In addition, these data clearly highlight the extent of Hex divergence in *Hydra* compared to that of *A. millepora*, and are consistent with *Hydra* belonging to the most diverged class of cnidarians (Bridge *et al.*, 1992 and 1995).

In spite of the potential for *D. melanogaster hex* to have a role in axis patterning similar to vertebrate *hex*, *in situ* hybridisation analysis (Figure 6.2) demonstrated that this was not the case. Except for a maternal deposition of this gene, *D. melanogaster hex* showed no early expression that could relate to an axis-specifying role similar to the anterior specification seen in vertebrates. Later in mouse embryogenesis, vertebrate *hex* is activated in the definitive endoderm. At least to a certain extent, parallels can be seen with the endodermal expression of *D. melanogaster hex*, where *hex* transcripts are confined to the region of the endoderm-derived posterior midgut primordium.

As the immunohistochemical analysis of the deficiency stock 3340 was unsuccessful, the significance of the possible mesoderm-specific *D. melanogaster hex* expression was unable to be ascertained. It was later discovered that the *UbxlacZ P*-element transgene employed to distinguish the heterozygous and homozygous deficiency stock embryos (see Section 6.2.3) was in fact a weak reporter, explaining the lack of specific staining seen during the analysis. Thus, whether the mesoderm expression is true, and whether it has any relationship to the mesodermal *hex* expression seen in the mouse, remains to be elucidated. Although not in the scope of this study, clarification of this issue would simply require the placement of the deficiency stock flies over a balancer chromosome incorporating a much stronger, and thus more easily detectable, reporter gene.

It is interesting to note that, on the basis of non-muscle myosin stains, it has been reported that the deficiency stock 3340 has no midgut phenotype (Bilder and Scott, 1995). This would infer that Hex plays no significant role in gut morphogenesis. Instead, Hex would more likely be involved in the specification and/or differentiation of cells within the gut. Consistent with this, in the midgut, *labial* (*lab*), which shows a similar expression pattern to *D. melanogaster hex* (compare Figures 6.2 and 6.4), has been demonstrated to play a role in the specification of endodermal cells (Hoppler and Bienz, 1994). Further, similar to the deficiency fly stock 3340, *lab* mutants have no morphogenic midgut phenotype. It is worth speculating at this point that this possible specification of *D. melanogaster* endodermal cells by Hex may parallel the role Hex plays in cell specification and differentiation, as reported in vertebrate endothelial and hematopoietic cells (Thomas *et al.*, 1998).

The early endoderm-specific expression of *hex* in higher metazoans predicted that *hex-Am* would be activated at an early stage of *A. millepora* development, possibly around the time of gastrulation and subsequent formation of the endoderm (24-hour-old embryos; see Section 1.4.1). Northern analysis proved this incorrect. Expression of *hex-Am* was initially detected in 48-hour-old embryos, and expression increased in intensity at the pre-settlement stage of development (Figure 6.8). Thus, it seems that *hex-Am* has no early role in either axis or tissue layer specification. Interestingly, the oral pore first appears after the completion of gastrulation, at approximately 48 hours post-fertilisation. Thus, it is possible to speculate that Hex-*Am* is involved in specifying the differentiation of endodermal cells into specialised cells of the oral pore, similar to the specification of cells seen in vertebrates, and possibly *D. melanogaster*. To confirm this, *in situ* hybridisation was performed on coral

embryos, using *hex-Am* as a probe. Although endodermal *hex* expression was detected, it was also present in the negative control. As described in Section 4.4, non-specific endodermal staining has consistently been reported as an artefact of experimental procedures performed on coral embryos serving to make these results highly ambiguous. Further attempts and modifications of the *in situ* hybridisation procedures, which are not in the scope of this study, will either confirm or disprove the above speculations.

In conclusion, this chapter reports the identification of both a *D. melanogaster* and an *A. millepora hex* homolog. The endoderm-specific expression pattern of *D. melanogaster hex* parallels that of its vertebrate homologs. Transcripts of *hex-Am* are detected during development around the time of oral pore opening and thus, similar to vertebrates, and possibly *D. melanogaster*, Hex-*Am* may play a role in the specification of the endodermal cells to their specialised fates. If so, *hex-Am* could potentially act as a marker for the differentiation of endodermal cells to a more specified fate.

7. Discussion

7.1 Introduction

Although a relatively new metazoan model organism, *A. millepora* has already provided valuable information to the community of molecular, evolutionary and developmental biologists. The aim of this thesis was to extend this current information to provide further insight into the evolution of metazoan development. Two approaches were employed. Firstly, a specific molecular analysis of the *A. millepora* DPP/BMP2/4 signalling pathway was performed. Secondly, an EST study was performed, opening way to a broader analysis of *A. millepora* molecular development.

7.2 The DPP/BMP2/4 signalling pathway

D. melanogaster DPP, and its vertebrate homologs BMP2 and 4, play central roles in the dorsal/ventral patterning of *D. melanogaster* and vertebrate embryos, respectively. Several facts have led to the assumption that these proteins evolved in these animals in the context of this function. Firstly, although a number of sponge TGF-β receptors have recently been isolated (Suga *et al.*, 1999), all were highly diverged members of the TGF-β- and activin-specific classes, with no similarity to DPP-specific receptors. Further, *C. elegans* has a highly diverged DPP signalling pathway (see Figures 4.2 and 4.5) and no true *C. elegans* DPP homolog has been isolated. However, before the commencement of this thesis, *dpp* was isolated in *A. millepora* (*dpp-Am*; Hayward *et al.*, submitted), an animal without an overt bilateral axis. Analysing DPP in this primitive radial animal would help decipher DPP's ancestral role and hopefully provide insight into the much-disputed origin of the dorsal/ventral axis. Thus, the object of this study was to examine DPP-*Am*, and characterise the DPP-*Am* signalling pathway.

The partial characterisation of the *A. millepora dpp* genomic locus (see Chapter 3) revealed, at least for the coding exons, conservation of the intron/exon organisation with both the *D. melanogaster dpp* and vertebrate *bmp2/4* loci. *D. melanogaster* DPP and vertebrate BMP2/4 are both expressed in a complex pattern throughout development. This expression pattern is controlled by a number of enhancer regions within the *dpp/bmp2/4* loci and as a result, several transcript structures of both *dpp* and *bmp2/4* have been detected (St. Johnston et al., 1990; Kurihara *et al.*, 1993; Feng *et al.*, 1995; see Section 1.9). Only two *dpp-Am* transcripts have been identified

(Hayward *et al.*, submitted; see Section 1.10), suggestive of simpler transcriptional regulation that may reflect a more ancestral level of complexity. *A. millepora dpp* enhancer regions are, to date, uncharacterised. It is now important to identify the genomic location of the 5' exon sequence of both *dpp-Am* transcripts, along with their enhancer regions. Comparative sequence analysis may then provide insight not only into the evolution of transcriptional regulation of this complex genomic locus, but also into whether specific exon and/or intron sequences have been lost or gained during the course of evolution.

To determine whether the sequence conservation of *dpp-Am* represented functional conservation, *dpp-Am* was ectopically expressed in *D. melanogaster* (see Chapter 3). Ectopic expression of either *dpp-Am* or *D. melanogaster dpp* (*dpp-Dm*), in both the *D. melanogaster* embryo and wing, led to similar phenotypes. This conservation of biological activity provides evidence that DPP-*Am* is a functional homolog of DPP-*Dm*.

The conservation of the *A. millepora dpp* genomic organisation, in conjunction with the conservation of DPP-*Am* biological activity, was suggestive of an essential ancestral role for this protein that has been maintained throughout evolution. In support of this, sequence characterisation of downstream components of the DPP-*Am* signalling pathway, specifically a DPP-specific serine/threonine kinase receptor and two DPP-specific intracellular Smads (see Chapter 4), demonstrated that they too were highly conserved. Although direct interaction between DPP-*Am* and these downstream components has not yet been demonstrated, transcripts of all these genes are present at the same *A. millepora* developmental stage (see Section 4.4). Further, there is a peak in expression of both the *dpp* and *smad* transcripts in the 24-hour-old embryo. It is interesting to note that, consistent with the work described in Chapter 5, all members of the DPP-*Am* signalling pathway so far isolated show higher sequence similarity to their vertebrate homologs than they do to their *D. melanogaster* homologs.

Expression analysis of *dpp-Am* (Dave Hayward and Eldon Ball; see Section 4.4) illustrated a confinement of transcripts to one end of the gastrulating *A. millepora* embryo. Previously it was assumed that the oral/aboral axis of cnidarians corresponded to the anterior/posterior axis of higher metazoans, especially with the advent of the zootype theory (Slack *et al.*, 1993; see Section 1.3.1). Recently, however, several lines of evidence have challenged this theory (see, for example,

Hobmayer *et al.*, 2000, Gauchat *et al.*, 2000, Section 1.3.1). Even so, there is no evidence of a correspondence between the dorsal/ventral and oral/aboral axis. The results reported here implicate DPP as functioning in axis specification, or at least gives DPP a role that could have evolved into an axis-specifying role. If the former is correct, it would suggest a more ancient origin for a DPP-specified axis than originally thought. However, it is important to note that it would still not explain why *C. elegans*, an overtly bilateral higher metazoan, has no DPP homolog. If DPP-*Am* plays no role in *A. millepora* axis specification, then alternative conserved functions could include, for example, an anti-neurogenic role (see Section 3.4.2).

It was hoped that *in situ* hybridisation and immunohistochemical analysis of the downstream components of the DPP-*Am* signalling pathway would provide further information to help elucidate the role of DPP-*Am*. The failure of these experiments highlights the problems associated with studying an organism where experimental procedures are only beginning to be characterised. However, protocols are continually being modified to achieve improved results, and soon not only will the expression patterns of the genes isolated in this thesis be examined, but, hopefully, so will the expression patterns of other genes known to be essential for higher metazoan axis specification (for example, see Section 5.4.2).

It is imperative when trying to decipher the role of a gene, not only to analyse the expression pattern, but also to characterise the function of the gene. Although at present no such work has been successfully performed in the coral, during this study three approaches to functionally analyse DPP were attempted. This included both loss-of-function experiments, via RNA interference (for example, Fire *et al.*, 1998) and morpholino oligonucleotides (Summerton and Weller, 1997), and a gain-of-function experiment, via ectopic DPP-*Am* expression. In all three procedures, 14-hour-old *A. millepora* embryos were selected, and incubated for several hours with either double-stranded *dpp-Am* RNA, a *dpp-Am*-specific morpholino, or DPP-*Am*. Although embryos initially survived, all embryos, including untreated control embryos, died at approximately 72 hours post-fertilisation. At least one reason for this has already been established. During the experiment the embryos were incubated in too low a volume of sea water. However, as functional experiments on coral embryos are confined to just a few days a year, during the annual mass spawning event, it is particularly difficult to make headway in analysing gene function. As the annual mass spawning event occurs at slightly different time points along the Queensland coast, one way to overcome this problem is to travel to various reefs over a period of time.

This, however, is costly. Alternatively, attempting to "culture" coral in a laboratory, so that it can be successfully manipulated to spawn on demand, should be seriously considered.

7.3 EST analysis

The aim of the second part of this study was to make use of the possibilities now open to molecular biologists to pursue a broader study of *A. millepora* molecular development. This encompassed an EST analysis of 3000 *A. millepora* cDNA sequences (Chapter 5). The result of this study was the identification of 509 ESTs that significantly matched sequences within the databases. These sequences encompassed a wide variety of gene functions and this fact, along with the surprising diversity of *A. millepora* genes already identified (reviewed by Ball *et al.*, in press), suggests that the *A. millepora* gene set is more complex than originally suspected.

Especially interesting are the developmentally important signalling/regulatory proteins and transcription factors that were identified in this analysis (see Section 5.2.2 and Appendix A.4). The most interesting ESTs will now need to be analysed individually. This will include complete sequence analysis, expression analysis and, hopefully soon, functional analysis. This has already been initiated with *A. millepora hex* (see Chapter 6). In vertebrates Hex is important in anterior specification (Thomas *et al.*, 1998; see Section 6.4), and is expressed in both the anterior endoderm (Tam and Steiner, 1999; see Section 6.4) and endodermally-derived tissues (for example, Martinez Barbera *et al.*, 2000; Pellizzari *et al.*, 2000; see Section 6.4). Thus, Hex was analysed as a potential *A. millepora* tissue, and possibly axis, marker gene (see Chapter 6). The results suggested that, as no early *A. millepora hex* (*hex-Am*) expression was detected, this gene does not function in early axis specification (see Section 6.3.3). However, *hex-Am* expression coincided with oral pore formation, and consistent with other analyses in vertebrates (see above) and *D. melanogaster* (see Section 6.2), this suggests a possible role for Hex-*Am* in the specification of these endodermally-derived cells. Future spatial expression analysis of *hex-Am*, although not possible in this study (see Sections 6.3.4 and 7.2), will help confirm the function of Hex-*Am*.

Analysis of all the ESTs with significant matches to known proteins showed that, although you would expect a similar conservation of gene sequence between coral and *D. melanogaster* compared to coral and vertebrates, in general coral genes match significantly better to their chordate homologs than to *D. melanogaster* or *C. elegans*

genes. Further, in the past it has been assumed that vertebrate-specific genes, where no protostome homologs have been identified, have recent origins. However, this study has demonstrated that genes previously thought to be vertebrate-specific are present in *A. millepora.* Thus, gene loss appears to have been more extensive in both *D. melanogaster* and *C. elegans* than was previously suspected, and many more genes central to higher animal development are likely to predate the Cnidaria/higher Metazoa split than initially thought. This suggests that *A. millepora* has inherited a common set of genes and patterning processes that were present in the common metazoan ancestor, but has perhaps not explored the developmental possibilities that these genes offer. In contrast, through co-option, gene duplications, and changes in both regulatory and coding sequences, higher animals have more fully exploited the range of evolutionary options offered by these genes. This re-confirms the ability of *A. millepora* to provide unique perspectives on the common molecular principles of animal development.

It is hoped that, in the future, these ESTs will be used to initiate a microarray project. This would allow developmentally important *A. millepora* regulatory genes to not only be identified via sequence matches to known genes, but also to be identified on the basis of gene expression patterns in the *A. millepora* embryo, larva and adult.

7.4 Summary

The comparative analysis of a variety of different animals is crucial to the study of the evolution of metazoan development. However, a significant proportion of scientific study focuses on the analysis of higher metazoans. Until recently, primitive metazoans that can represent an outgroup for these studies have been lacking. The primitive cnidarian *A. millepora* has many attributes that make it an excellent candidate for these evolutionary studies. However, as a relatively new metazoan model organism, there is a limited knowledge of this animal. The initiation of an EST analysis performed during this study has provided the data needed to accelerate information available regarding *A. millepora* molecular development. In addition, a more specific analysis of the *A. millepora* DPP signalling pathway, has provided valuable information regarding both the ancestral function of this transduction cascade, and the origin of the higher metazoan bilateral axes. The results of the studies reported here have clearly illustrated that *A. millepora* is, indeed, a great system to work with, and hopefully a system that will continue to provide a wealth of information to help in the pursuit of understanding the evolution of animal development.

Appendix

Appendix A.1 Complete nucleotide sequence of Bmpr1-*Am* cDNA.

GCACGAGCAAATTCCAAGACGATATACGACCCCTGTCTGGGGCCAACGTAACGACTTTGTGTAATGCTTCATGAT
CTTCAGCACTCGGTGGGTTGAAAATCCGTAATTGCAAACAGTGTCGATTGGCGTGCAGAGTCTTTCGATAGCTCT
CAAAGATTTATGCTAGGCCTCTCTGCTCGGTAGGGGGTGAAAAAACGTCACTTCTCCTTTTGCGGCAAGTCGTGA
AAAAATTCTATGGATGACCTTAATCTCCGTTGAGCAATATGGCAAAACCGACTATGGCAAGTGTGTCCCTGCTTC
ACTTGGTAGTTCTTCTCTGCTTTCTGGCTCATGGTACTAAGAGCATACGATGCAAGTGTAGCACCCACAATTGCC
CAGGGGACCACATCAATGAAACGTGTACAACAGAAGGTCACTGTTACAAGAAAGTGGAACAAAGTGAAGAGGATG
GTCTAGAGTATGTTACATATGGCTGCCTTCCTCCTGAAGAGCAAACCACAATGCAGTGCAAAACACCAAATCACA
TTCACACCAGACTGCTCTCTATAGAATGTTGTAGTAAAGACCTGTGCAACGATGTTTTACAACCTAAGCTACCGA
CAACAGCACCCCCCACAACAATTACAACGGTTGAAGAGGAGACAGAAGAAGCTGTCACAGAACAATATTCTATCC
TCTTCATTAGCGCAGGTGTCTGTGTAGCAGTCTTTGTGATTTTCCTTGGAGTCCTTTGTTGCCGATTGAGAGCGA
CACGCAGCAGACTTCCCTTTCCCTTTGAAGTGGAGAAGTATGGCAGCCCTTATATGTCTTCAGGGGAAACACTCA
AAGACATGCTGGATCAAAGTTCTGGGAGTGGCTCAGGATTACCACTGCTGGTTCAAAGAACCATTGCTAAGCAAG
TGACGTTGGTAAGAAGTGTGGGTAAAGGCAGATATGGTGAAGTGTGGCAAGCAAGATGGCGAGGAGAGGACGTGG
CTGTCAAAATATTCCTGTCACATTGTGAATCCTCATGGCAGAGAGAAACTGAGATCTACCAGACCGTTTTACTGC
GGCATGAGAGCATTCTAGGCTTCATAGCATCAGACATTATTGGAAGCAATCAAGTGACACAGATGTATCTCATAA
CAGATTATCATCCTTATGGATCATTGTATGATTTCTTACGATGCCACTGCCTCAACAAGAAGACTATGATAAGGC
TTGTGTTGTCTGCATCAGCAGGCTTGACGCATCTTCATACTGAAATCCAGGGGACAAAAGGAAAGCCTCCTATAG
CTCATCGTGACATGAAAAGCAAGAACATCCTTGTCAAAGAGAACTTGACCTGCTGTATAGCAGATTTTGGACTTG
CAGTGAAGTACTCGCCAGAAACTGAAGAGGTAGACATCAAGCCAGACACAAGAGTGGGAACACGGCGATACATGG
CCCCTGAAGTTCTTGACAATGCGTTGGATTCAAGGAACTTTGCTGCTTTTAAGATGGCAGATATGTACTCATTTG
GATTAGTGTTATGGGAGATTGCCCGAAGGTGTTTTACGGATGAAACTGGACTGTGTGAGGAGTACCAGATTCCTT
ACTACGATATGCTTCCTGGAGATCCTTCTTTTGATGAAGTCAAAAGAGTTGTGTTGACAGACAAGAGAAGACCCT
CAGTGCCAAATAGATGGTACAGAGATGAGTGTCTCCAAACTATGGCCAAGCTGATGACAGAGTGTTGGGCACAAC
ACCCTGCAGCCCGTCTGACAGCCTTGAGAGTTCAGAAAACTTTAAGCAAACTCAAGAAGTCAATGGATTTCATAG
ACCAACCATATGATGCAGACAATGATAGCCCCAGGACAAGTGTCACCACAGCCTAAGAGGCATACTGCTTGAATC
ATTCATCAGAAACAAAGCTACCAGTCAGTTGTATTTGAGACTCATTTATTTTGTAGATGATGCTGTTTCTTAGCA
ATGCAAAATTAATCATCTCTGAAGGATTTAAGGAAAAATTTACGCTGTTGGTAATGGGGAATTGCTTGCCACATC
AGACTTCAGAAAAGATCAAAGAAAGGGCAAATACACCTTTAAAAGCTCCAAACTGTTTGCACAAACCGCTATAAT
AATGTATGCCCCAGTCTCAGGGAACTTGTTGGATAGGCTGTGATAAATATTCAATATCATTCATTAAATATGGAG
AAAGTATAGAACAGTTGAAATTTCAGCCAAGTAAGGACAAGGTTTGCAACTTTGTACAAAGAACTACTGCATTGA
AGTATCTGCCTTTGCATGCACAGTAAATGAAACCCGTGGGCACTTTGCCTTCAACATACAGCTTGTACACGTACG
TGCAGTCAAACCCCAGTACTGTGCTATGGAGCTGTGGGGATTGAGAAAAATTGTCAAAGTTCTGGATTGTTGAAA
AAATGAAAACAAAGGAAAAAGCAGGAGAGAAATACAAAAATAATGGTCACAGAGTCAAGTAACAGCAGCAAGAAA
AAAGTGATATTCATGGTAATTTACTTCACTGCCCACAGCATATGTTAGACTGGTTATACTTTTAGCTCTTGTAAG
GGAATGCGCAGTTATTTCCACTGCACATTCAGAGTTGTAAGAACAATTATTTAGAGGGAGAAGTTACATAATTTT
TTTCCTCTGCTCTCATAAATTTTCTCGCTCTACTGGGAAATTTTTCATGACAAGAACTAATATACCACTAGCCAG
TGCTGCTACCAGTTGCCAGTGCATTTTGAGAACCCTTCTTATTGACATAGCTTAGTAACCAAATCAACATCTTTG
GTCATTGATATGAAAGTCTAAATGAGTAATCTTCAACCATATATAATGCAGTTTTTAGCTGCAATAATACTTTTT
AAATATTGTAAAGGGATGGAGAATTGCAAACAAATCAACCTGGTTGCATTTTGATTCCTTGTCTTATTTGGTGGA
AAAGAGAAAATGAGTTGTAATATTAGGAAAAACTGTATCGAAATTGAAGATTAGGAAAAGATCAAAATGTTGTTA
TCAAGGAAAATGTTTTGTTTGCTATAGAAAGATAGCAAGATTCAGAACTGTGCAACAGTAATGAGAATGAATTTT
GTTTGAATGTGTGCATGCTTTGTTCTGATCAAAAGTGTGCCAAATTTTTCTTGGAGTTTCATTAAATTCTTTTCA
GTCCCGACTACTATAATTTTAAAAAAAAAAAAAAAAAAAA

Appendix A.2 Complete nucleotide sequence of Smad1/5a-*Am* cDNA.

TTTCGGCACGAGAGCACTTGTGTTTTATGATCAAGCACGCCCATCAATTCGCCATTTGAAGTCGTTAACTTCTCT
GAAACGATTCTAGGACGTTGTAAGTCAGGTTTCTCAAGAAAGGACTGTTAATTCTGTTCCTCCCAACCAACACCG
CGCGCAGACAACCTTCCGAAGGGAGACTTTGCAAATCAAGATGTCAAATATGTCTTCCTTGTTTTCTTTCACTCA
CCCAGCCGTAAAAAGACTTCTAGGCTGGAAACAGGGTGACGAAGAAGAAAAATGGGCGGAAAATGCTATCGGTTA
TTTATCTAAGAAGCTCAAGAAAAAGAAAGGCGCCTTGGAAGAATTGGAGAAAGCGTTATCAAACCCTGGCCAGCC
GTCAAAATGTGTAACAATAGCCCGTAGTTTGGACGGGCGTATTCAAATACTCCATAGAAAAGGTTTACCGCATGT
AATTTACTGCCGTGTGTGGCGATGGCCTGATCTTCAAAGTCATCATGAACTAAAGCCATTGGATTGTTGCGAGTA
CCCTTTCGGAATGCCGAAACAGAAAGAAATCTGTATAAATCCTTACCACTACAGACGGGTGGAAAGTCCGGTTTT
GCCACCTGTATTGGTTCCCAGACCGAGTAATTCGTGCCCAAAACCACCTCCACCTTTGCCATCTCCCTTCAAAAG
CACAGATGAGCCACCAATGCCTTACAATGCTTGTTTCCCCTTCAAATCAACTCAGCAACCTTCTCCTAGCCTGAG
CCCGGCAATGTATATTGCAGAATCACCTAAAAGTTACATTTCCGAGGAAGGTGGCTCTCCATGTTCCATGGACCC
TAATGCTATGGATATAGATGCCAACGGCACCTTGCCGACAATGGTGAACAACCAGCCAGGATATCTTTCACATGT
GACACCAGTTAATTACCAGGAGCCGTCATCATGGTGTTATGTGTCATACTACGAACTCAACAACAGAATTGGTGA
TCGGTTCTATGCTAACAGCACCAGTATTATTGTTGATGGCTTCACAGACCCTAACACTGGAAACTCTGAAAGATT
CTGCTTGGGTTTGCTATCCAATGTTAACCGTAACTCGACCATCGAGAGCACTATAAGGCACATTGGGAAAGGTGT
ACACTTGTATTACGTTGGAGGTGAGGTTTATGCTGAATGCCTCAGTGACTCAGCCATCTTTGTTCAGAGTCGAAA
TTGCAACCACAGTCATGGTTTCCACCCAACAACAGTCTGTAAGATTCCCTCAGGATGCACGCTTAAAATTTTCAA
CAATCAAGAATTTGCACAGCTGCTTTCACAATCCGTTAATTATGATTTCAAGGCAGTGTATGAGTTGATTAAGAT
GTGTACAATTCGTATGAGTTTTGTCAAAGGCTGGGGCGCAGAGTATCATCGTCAAGATGTCACCAGTACACCCTG
TTGGATTGAAATCCACCTCCTCGGCCCCCTTCAGTGGCTGGATAAGGTGCTTTACCAGATGACCATGCCAGAAAA
TGGGCCTACCTCTCACTCAGCATAAAGTTGGACTTTTTATAAATGAATGTTAAGTTATTGCCTCTTGCAATTGAA
AAAGCTGTTGTTATTTAACCTCATGAAGAATTATTCATAAAGGAAGGACATCTTTCACCTCCCTTTAAAAAGCCT
TGTTTGTTTGTCGGTCTTTGTAAAAACGTCTTTTATTCCAAAAAGAAACAATAATAATATTGTCCTGACTTTTGT
ATCAGTGTTAAATTTTACAAACTTGTCTTTGCCTTTAGTTTAATATTATTGCTATTGGACAACCTGTCACAATGA
TTTTGGTGGTACTTTTCTGGAATAAATGGGTAGAGAGGGTTGAAGAAAGTTTTTTTTTTTTAATCTTATAATTTG
TTTTACGTTGTTACTATTTTTAATAAAAAAAAAAA

Appendix A.3 Complete nucleotide sequence of Smad1/5b-*Am* cDNA.

GGCACGAGGTTACTATCTCTCTTAAGCCTTAAATTACGACGTCCTCGACAATAATCGGAACATTAGCGTGTGCCT
AAACCAAACCTGGGAGACAAGCTTCCTGTAAAAGTGAAGAGGACTTAAAAACTGACTGCCAATATGACGACTATG
GCTTCGTTGTTCTCGTTTACTCCGCCTGCGGTGAAGAGACTTCTCGGCTGGAAGCAAGGAGACGAAGAAGAAAAG
TGGGCGGAAAAAGCTATTGAGTCCTTAGTGAAGAAACTCAAAAAGAAAAAGGGTGCACTGGAAGAGCTGGAAAAC
GCACTTTCAAAACCGGGTCAGCCGTCGAAATGTGTCACTATTACACGTAGTTTAGACGGCCGTCTGCAGGTCTCG
CACAGAAAAGGTTTGCCTCATGTGATTTACTGCCGAGTATGGCGATGGCCCGATCTTCAAAGCCACCACGAGTTA
AAGCCATTGGATTGTTGTGAATATGCGTTTGGATTGAAACAAAAAGAAGTTTGCATTAACCCTTACCACTACAGG
CGTGTGGAAAGTCCGGTTTTACCCCCGGTCCTTGTGCCAAGACAAAGTGAATACCCAAGACCTCCACCGCATCTA
CCTTTGCCATTCAGAAGCGCTGAAGATCCACCAATGCCTTATAATGCCTCATATCCTTTCTCAAATCGTCCAAAT
CAGCTGGAACACAGCCCAGCAACCAGTTTTGAAATGCCAGAGACACCTGCAGGGTATATATCTGAGGATGGAGGG
TCTCCACGACCAGTGGACCCAAATGCAATGGATGTAGATACTGTTAATTCTCCCCCTGCAGTAACGCCACAGACT
GAGTTATCTCACGTCACTCCAATCAACTACCAGGAACCATCCACATGGTGTTCTGTTTCTTATTACGAGCTTAAC
AATCGCGTTGGACAACAGTTCCAAGCCCACAGCACAAGCATTTTTGTTGATGGCTTCACAGACCCTAACACTGAG
AACTCTGAAAGATTCTGCTTGGGTTTGCTATCCAATGTTAACCGTAACTCAACCATTGAGAGCACAAGAAGGCAC
ATTGGAAAAGGTGTACACTTGTATTATGTTGGAGGCGAAGTTTACGCTGAATGCCTTAGTGACTCAGCCATCTTT
GTTCAGAGTCGAAATTGCAACCACAGTCATGGTTTCCACCCAACAACAGTCTGTAAGATTCCTTCAGGATGCACG
CTTAAAATTTTCAACAATCAAGAATTTGCACAGCTGCTTTCACAATCTGTTAATTATGGTTATGAAGCTGTGTAC
GAGTTAACAAAGATGTGTTCAATCCGTCTGAGTTTTGTCAAAGGCTGGGGTGCGGAGTATCATCGTCAAGACGTG
ACAAGTACACCATGCTGGATTGAAATTCATCTTCATGGTCCTCTACACTGGCTTGACAAGGTGTTAAGTCAAATG
GGTTCCCCAAGAAATCCAATATCTTCTGTCTCTTAAACCAGTCATTGCATGCATTTAAGTTTTTAATCAGCATTG
GGGTAACTCTGTCACTTTTTTTTATAAAATTGCATTTGATTGGAACATGGGACTTGCAAATAGGGAGCATGCTTT
AGGTTTGGAGAAAAGATAGCTGCCTAAAGGCTGTTCAAGCTATTGAGTGGAAATGCTTTGTTAACAGATGAACTG
CAGATTTGCATGAAAGAAATATTTTCCACTGATTTGAGGCATTTCTAGAGCAACATTCTGAATTTTTTTTAACAC
CATTTGGCTCATTTTCCCTTCCCACTTTTGATTCATTAGCACAAGGTAAAGCACAATTTGTTTCAAATCTCAAAA
ATGCTAAGGTATCGGGACTGCTTGAATGTAATATTATTGTAATTTTATGCCACAAATTATACAACACATTTCATT
GTGTCACATAATTTTTGTACTGTTTGACATTTATATAAATATTACCTTCTCATAGATTGATTCCTAAAATTTAAA
AAAAAAAAAAAAAAAA

Appendix A.4 EST sequences classified into functional classes.

Column 1: Clone position (Plate and well number in library)

Column 2: Genbank accession number of most similar sequence to clone

Column 3: Gene description of most similar sequence to clone

Column 4: Probability of sequence match occurring by chance (E value)

A. PROTEINS INVOLVED IN DNA STRUCTURE, REPLICATION AND REPAIR

General:

P A02	Q9DC33	mus musculus (mouse). 1200004e06rik protein	5e-18
1 F01	O75531	homo sapiens (human). barrier-to-autointegration factor	2e-23
A E03	Q9H3J0	homo sapiens (human). brain my036 protein	4e-41
V H10	P54278	homo sapiens (human). pms1 protein homolog 2	1e-16
J D02	P54727	homo sapiens (human). uv excision repair protein protein rad23 homolog b (p58)	3e-24
J E02	P54726	mus musculus (mouse). uv excision repair protein protein rad23 homolog a	2e-10
K A06	Q9DAA6	mus musculus (mouse). 2610104c07rik protein	5e-45
D C05	Q9W256	drosophila melanogaster (fruit fly). cg11301 protein	2e-17
F GE4	O35651	mus musculus (mouse). dhm2 protein	2e-42
S B12	Q99J62	mus musculus (mouse). similar to replication factor c(activator 1) 4 (37kd)	5e-25
J D08	P55862	xenopus laevis. dna replication licensing factor mcm5 (cdc46 homolog)	9e-19

Histones/chromatin associated factors:

H A06	Q04841	mus musculus. dna-3-methyladenine glycosylase	7e-17
H D02	P07796	strongylocentrotus purpuratus (purple sea urchin). histone h1-gamma, late	7e-16
C C08	P08991	strongylocentrotus purpuratus (purple sea urchin). histone h2a variant (fragment)	2e-50
U E02	P08903	encephalartos altensteinii (altenstein's bread tree) (cycad). histone h3	7e-32
Z D12	Q9UBB5	homo sapiens (human). methyl-cpg binding protein 2	6e-12

B. TRANSCRIPTION FACTORS AND ASSOCIATED PROTEINS

D D08	Q9V868	drosophila melanogaster (fruit fly). cg10933 protein	4e-13
B B03	Q99581	homo sapiens (human). fev protein	2e-21
B D09	Q9VW37	drosophila melanogaster (fruit fly). hlh106 protein	1e-18
G H01	Q12772	homo sapiens (human). sterol regulatory element binding protein-2	1e-25
E B05	Q9DFZ4	xenopus laevis (african clawed frog). churchill protein	5e-29
H D01	Q9DFB3	xenopus laevis (african clawed frog). hand2' (fragment)	3e-17
K A11	Q23819	chlorohydra viridissima (hydra). homeobox gene cnox3 mrna (fragment)	5e-18
E E10	Q9QYF5	mus musculus (mouse). pancreas transcription factor1 p48 subunit	9e-38
B D02	Q9U9U4	drosophila melanogaster (fruit fly). hey (cg11194 protein)	1e-17
C G12	Q15910	homo sapiens (human). enhancer of zeste homolog 2 (enx-1)	4e-50
N H12	Q9VE04	drosophila melanogaster (fruit fly). cg14283 (putative transcription factor)	2e-08
U H05	O13023	xenopus laevis (african clawed frog). homeodomain protein xhex	5e-32
V C03	Q9BMZ8	drosophila melanogaster (fruit fly). af10	6e-21
P D03	P21573	xenopus laevis (african clawed frog). y box binding protein-1	1e-40
E G01	P35269	homo sapiens (human). transcription initiation factor iif, alpha subunit	2e-16
W C04	Q9H8W5	homo sapiens (human). cdna flj13181 fis, clone nt2rp3004016	1e-16
C B09	Q13573	homo sapiens (human). nuclear protein skip	2e-65
Y D01	Q9YGM0	gallus gallus (chicken). mll protein (fragment)	1e-45
R E05	Q9W1H0	drosophila melanogaster (fruit fly). cg5591 protein	3e-31
E A07	Q9CZL5	mus musculus (mouse). 2700061n24rik protein	1e-34
U G11	O76071	homo sapiens (human). wd40-repeat containing protein ciao 1	1e-17
K E12	Q9Y6J9	homo sapiens (human). pcaf associated factor 65 alpha	2e-14
V H11	Q14839	homo sapiens (human). chromodomain helicase-dna-binding protein 4 (chd-4)	8e-26
B C03	Q24423	drosophila melanogaster (fruit fly). noc protein	2e-07
O D12	O77161	branchiostoma floridae (florida lancelet) (amphioxus). snail	1e-38
G C08	P70670	mus musculus (mouse). nascent polypeptide-associated complex alpha polypeptide (muscle-specific form gp220)	1e-09

C. RNA-BINDING PROTEINS AND PROTEINS INVOLVED IN SPLICING

H F01	Q9XIK2	arabidopsis thaliana (mouse-ear cress). t10o24.12	6e-10
H C07	Q9U239	caenorhabditis elegans. y56a3a.18 protein	8e-21
H G03	Q9Y5A8	homo sapiens (human). ny-ren-6 antigen (fragment)	2e-30
K D12	Q9VVU6	drosophila melanogaster (fruit fly). cg6841 protein	8e-31
H 12	Q9GQI7	caenorhabditis elegans. sf2	5e-29
T D01	Q9VGW8	drosophila melanogaster (fruit fly). rbp1 protein	5e-18

H G9	Q9VFT0	drosophila melanogaster (fruit fly). b52 protein	1e-24
V C04	Q9DAW6	mus musculus (mouse). 1600015h11rik protein	4e-22
1 F01	Q9D6C6	mus musculus (mouse). 3632413f13rik protein	8e-21
Q B08	Q9UK45	homo sapiens (human). u6 snrna-associated sm-like protein lsm7	2e-37
Z C06	Q9DBT2	mus musculus (mouse). 1200014h24rik protein	5e-44
C D02	Q23796	chironomus tentans (midge). hnrnp protein	2e-25
A D05	P43331	homo sapiens (human). small nuclear ribonucleoprotein sm d3	6e-22
T G11	P48810	drosophila melanogaster (fruit fly). heterogeneous nuclear ribonucleoprotein 87f	2e-18
J H10	Q9SP10	medicago sativa (alfalfa). glycine-rich rna binding protein (fragment)	2e-17
A A10	O88532	mus musculus (mouse). zinc finger rna binding protein	2e-10
P H07	O89086	mus musculus (mouse). putative rna-binding protein 3	2e-14
W C10	Q9D9V1	mus musculus (mouse). 4921506i22rik protein	1e-13

D. PROTEINS INVOLVED IN TRANSLATION

Ribosomal proteins:

E A01	Q9CZI5	mus musculus (mouse). 2410044j15rik protein	7e-56
O B01	Q9CQK2	mus musculus (mouse). ribosomal protein s24	1e-53
D B02	Q9CZM2	mus musculus (mouse). ribosomal protein l15	1e-34
K C01	P50886	xenopus laevis (african clawed frog). 60s ribosomal protein 122	6e-15
I C09	P25112	homo sapiens (human). 40S ribosomal protein s28	1e-18
Q C09	P23358	rattus norvegicus (rat). 60s ribosomal protein l12	3e-68
K D8	Q9DC49	mus musculus (mouse). repeat family 3 gene	3e-31
Q D08	P17078	rattus norvegicus (rat). 60s ribosomal protein l35	5e-33
E H09	Q9BGU2	sus scrofa (pig). 60s ribosomal protein l35	3e-25
D D09	Q9VXX8	drosophila melanogaster (fruit fly). ribosomal protein l37	1e-40
O E02	P29316	homo sapiens (human), and rattus norvegicus (rat). 60s ribosomal protein l23a	2e-53
J E03	P23403	xenopus laevis (african clawed frog). 40s ribosomal protein s20 (s22)	9e-11
1 H01	P32969	homo sapiens (human). 60s ribosomal protein l9	1e-57
1 C08	Q9CSN9	mus musculus (mouse). 2410044j15rik protein (fragment)	1e-29
E H03	Q9GP16	heliothis virescens (noctuid moth) (owlet moth). ribosomal protein l31	6e-49
J A12	P50894	fugu rubripes (japanese pufferfish) (takifugu rubripes). 40s ribosomal protein s7	2e-59
K B11	Q9Y3U8	homo sapiens (human). 60s ribosomal protein l36	2e-16
H D11	P34737	podospora anserina. 40s ribosomal protein s15 (s12)	1e-45
H E04	P79016	schizosaccharomyces pombe (fission yeast). 40s ribosomal protein s19 (s16)	3e-18
O F05	P17702	rattus norvegicus (rat). 60s ribosomal protein l28	1e-26
J F12	P25232	homo sapiens (human). 40s ribosomal protein s18 (ke3)	1e-58
E G04.	Q9XS36	sus scrofa (pig). ribosomal protein l29/heparin/heparan sulfate interacting protein	2e-20
E G06	P04643	homo sapiens (human). 40s ribosomal protein s11	2e-53
Q G12	O61231	drosophila melanogaster (fruit fly).60s ribosomal protein l10	4e-67
K H04	Q9D1I6	mus musculus (mouse). 1110006i11rik protein	4e-13
A A12	P41125	gallus gallus (chicken).60s ribosomal protein l13 (breast basic conserved protein 1)	6e-49
B B04	P25111	homo sapiens (human). 40s ribosomal protein s25	1e-26
B D04	Q9VFB2	drosophila melanogaster (fruit fly). cg4247 protein	4e-26
B E01	Q98TR7	heteropneustes fossilis. ribosomal protein s16	9e-54
B E03	Q9CQM8	mus musculus (mouse). ribosomal protein l21	4e-50
B G07	P02362	xenopus laevis (african clawed frog). 40s ribosomal protein s7 (s8)	1e-58
C C01	P39028	homo sapiens (human). 40s ribosomal protein s23	3e-64
C G09	Q9CXW7	mus musculus (mouse). 3010033p07rik protein	3e-55
F B08	Q9UNX3	homo sapiens (human). ribosomal protein l26 homolog	1e-42
F H08	P02433	homo sapiens (human).60s ribosomal protein l32	1e-42
G D08	Q9IA74	paralichthys olivaceus (flounder). 40s ribosomal protein s15a	3e-60
G F06	Q9BUZ2	homo sapiens (human). similar to ribosomal protein l34	1e-29
G H09	Q02546	homo sapiens (human). 40s ribosomal protein s13	6e-74
M A11	P23131	homo sapiens (human). 60s ribosomal protein l23 (l17)	4e-62
M C11	Q9DFQ7	gillichthys mirabilis (long-jawed mudsucker). 60s ribosomal protein l24	6e-43
M E02	P14148	mus musculus (mouse). 60s ribosomal protein l7	2e-40
M E03	P06584	cricetulus griseus (chinese hamster). 40s ribosomal protein s17	2e-55
N D11	Q9GNT2	stoichactis helianthus (carribean sea anemone). 60s ribosomal protein l27	6e-45
P B12	Q9VWG3	drosophila melanogaster (fruit fly). cg14206 protein	2e-38
S B09	O46160	lumbricus rubellus (humus earthworm). 60s ribosomal protein l14	1e-21
S E12	P23396	homo sapiens (human). 40s ribosomal protein s3	5e-19
T B03	P02401	rattus norvegicus (rat). 60s acidic ribosomal protein p2	5e-24
T D04	P13471	rattus norvegicus (rat). 40s ribosomal protein s14	3e-52
T H04	Q9IA75	paralichthys olivaceus (flounder). ribosomal protein l17	2e-65
U A03	P47830	xenopus laevis (african clawed frog). 60s ribosomal protein l27a (l22)	3e-51
U G08	P19944	rattus norvegicus (rat). 60s acidic ribosomal protein p1	3e-12
V C01	P49180	caenorhabditis elegans. probable 60s ribosomal protein l35a	9e-24

V C11	O16797	drosophila melanogaster (fruit fly). 60s ribosomal protein l3	7e-57
V E01	Q9VZD5	drosophila melanogaster probable mitochondrial 28s ribosomal protein s6	1e-16
V B01	P46405	sus scrofa (pig). 40s ribosomal protein s12	1e-47
V D02	Q9C8F7	arabidopsis thaliana (mouse-ear cress). 60s ribosomal protein l30, putative	9e-08
V G08	O76968	podocoryne carnea. ribosomal protein l18a (fragment)	4e-41
W F10	P35687	oryza sativa (rice). 40s ribosomal protein s21	2e-16
W G05	Q9CQF1	mus musculus (mouse). ribosomal protein l18	8e-61
1 H05	Q91487	salmo trutta (brown trout). 60s ribosomal protein l13a (transplantation antigen p198 homolog)	4e-40
Non-ribosomal:			
B H11	P70531	rattus norvegicus (rat).ca2+/calmodulin-dept. eukaryotic elongation factor-2 kinase	6e-19
V F05	P43896	bos taurus (bovine). mitochondrial elongation factor ts precursor (ef-ts) (ef-tsmt)	1e-11
F G11	Q9BVM0	homo sapiens (human). similar to block of proliferation 1 (fragment)	2e-13
D A01	Q9V859	drosophila melanogaster (fruit fly). cg4954 pr	4e-11
X B05	P70445	mus musculus (mouse). phas-ii	6e-24
O A09	Q9NR50	homo sapiens (human). eukaryotic translation initiation factor eif2b subunit 3	5e-23
V B09	Q91581	xenopus laevis (african clawed frog). polyadenylation factor 64 kda subunit (cstf)	7e-41
R E07	O00303	homo sapiens (human). eukaryotic translation initiation factor 3 subunit 5 (p47)	2e-14
T C06	Q9CRE9	mus musculus (mouse). eukaryotic translation initiation factor 2, subunit 3, structural gene x-linked (fragment)	2e-29
J G10	Q9BW18	homo sapiens (human). similar to cleavage + polyadenylation specific factor 6, 68kd subunit	4e-17

E. CYTOSKELETAL PROTEINS AND THEIR REGULATORS

Q D04	Q91955	gallus gallus (chicken). myotrophin (v-1) (granule cell differentiation protein)	1e-24
G A01	O96063	dugesia japonica (planarian). myosin heavy chain (fragment)	2e-16
A F03	Q9UKN7	homo sapiens (human). myosin xv (unconventional myosin 15)	6e-25
W D05	Q9CQ19	mus musculus (mouse). transient receptor protein 2	8e-42
U A08	P54357	drosophila melanogaster (fruit fly). myosin ii essential light chain (nonmuscle)	5e-11
X D04	Q9VUX6	drosophila melanogaster (fruit fly). cg5891 protein	1e-10
Z B06	P53462	heliocidaris erythrogramma (sea urchin). actin, cytoplasmic i	2e-65
S E07	O61377	dugesia polychroa. actin 3 (fragment)	1e-08
D B06	P53458	diphyllobothrium dendriticum (tapeworm). actin 5 (fragment)	9e-65
Q F04	Q9QXZ0	mus musculus (mouse). actin cross-linking family protein 7	3e-16
VE09	O15509	homo sapiens (human). arp2/3 complex 20 kda subunit	1e-71
Y D10	Q99M56	mus musculus (mouse). actin-like 6	5e-14
C D08	Q9W7F2	xenopus laevis (african clawed frog).wd-repeat protein 1(actin interacting protein 1)	8e-30
F A03	Q9U3S7	caenorhabditis elegans. b0272.5b protein (fragment)	8e-35
H B06	Q92747	homo sapiens (human). sop2-like protein	6e-56
1 A01	P18320	anthocidaris crassispina (sea urchin). profilin	1e-18
J C07	Q9NDD6	riftia pachyptila (tube worm). fibrillar collagen (fragment)	4e-27
O H09	Q9D320	mus musculus (mouse). 9030612i22rik protein	9e-46
K H03	Q00174	drosophila melanogaster (fruit fly). laminin alpha chain precursor	2e-17
K H09	O18342	halocynthia roretzi (sea squirt). beta-tubulin	4e-67
J D10	P18258	paracentrotus lividus (common sea urchin). tubulin alpha-1 chain	9e-73
A E08	P41383	patella vulgata (common limpet). tubulin alpha-2/alpha-4 chain	4e-57
D A05	Q9MZD2	oryctolagus cuniculus (rabbit). hypothetical 30.9 kda protein (fragment)	6e-24
E B12	P28738	mus musculus (mouse). neuronal kinesin heavy chain (nkhc)	1e-18
E B06	O88658	rattus norvegicus (rat). kinesin-like protein kif1b (fragment)	9e-43
B C09	Q9ULI4	homo sapiens (human). kiaa1236 protein (fragment)	4e-19
B G11	Q9U679	strongylocentrotus purpuratus (purple sea urchin).kinesin-c	3e-62
M D01	Q09997	caenorhabditis elegans. putative kinesin-like protein r144.1 in chromosome iii	2e-29
F A09	Q9HCG9	homo sapiens (human). kiaa1603 protein (fragment)	4e-49
J H12	P39057	anthocidaris crassispina (sea urchin). dynein beta chain, ciliary	5e-50
G G05	Q9JJ79	rattus norvegicus (rat). cytoplasmic dynein heavy chain	2e-63
U H11	Q9Y2F3	homo sapiens (human). kiaa0944 protein (fragment)	3e-55
S A11	O76108	anthocidaris crassispina (sea urchin). outer arm dynein lc3	4e-39
V C07	Q9D0M5	mus musculus (mouse). 6720463e02rik protein	7e-46
S D12	Q9U8W7	tachypleus tridentatus (japanese horseshoe crab). techylectin-5b	1e-15
Q E01	Q9WV92	mus musculus (mouse). protein 4.1b	1e-13
J B04	Q9JKK7	mus musculus (mouse). neural tropomodulin n-tmod	4e-11
H A1	P16086	rattus norvegicus (rat). spectrin alpha chain (fragment)	5e-18
K E09	P97814	mus musculus (mouse). pest phosphatase interacting protein	3e-25
H G08	P42768	homo sapiens (human). wiskott-aldrich syndrome protein (wasp)	1e-06
J G04	Q9UPY6	homo sapiens (human). wiskott-aldrich syndrome protein family member 3	6e-22
E A04	Q15142	homo sapiens (human). polycystic kidney disease-associated protein	5e-26
K H11	O88327	mus musculus (mouse). alpha-catenin related protein	8e-23

T B09	Q9VYR4	drosophila melanogaster (fruit fly). fw gene product	7e-12
V A01	Q25100	hydra magnipapillata (hydra). beta-catenin	5e-13
V D03	Q04615	xenopus laevis (african clawed frog). vinculin (fragment)	1e-15
X G02	Q9QWW1	mus musculus (mouse). homer-2b	8e-44
F B05	Q9I954	cyprinus carpio (common carp). thymosin beta b	2e-10
M C06	P26044	sus scrofa (pig). radixin (moesin b)	1e-36
S A01	Q9ET47	mus musculus (mouse). espin	2e-13
S D07	P42325	drosophila melanogaster (fruit fly). neurocalcin homolog	2e-73
E G04	Q9CS20	mus musculus (mouse). 5730465c04rik protein (fragment)	7e-11
U G01	P43689	biomphalaria glabrata (bloodfluke planorb). tropomyosin 2	2e-07
C A01	Q9H876	homo sapiens (human). cdna flj13898 fis, clone thyro1001738 (fragment)	1e-39
D D03	Q9BK91	strongylocentrotus purpuratus (purple sea urchin). dystrophin-likeprotein	5e-07

F. METABOLIC PROTEINS

General metabolic enzymes:

K A01	Q9N126	bos taurus (bovine). photoreceptor outer segment all-trans retinol dehydrogenase	5e-08
Q A08	Q9STP8	arabidopsis thaliana (mouse-ear cress). putative acyl-coa binding protein	2e-11
K B07	Q9QUJ7	mus musculus (mouse). long-chain-fatty-acid--coa ligase 4	1e-16
E B03	P53665	arabidopsis thaliana acyl carrier protein, mitochondrial precursor	5e-16
1 E04	P24368	rattus norvegicus (rat).peptidyl-prolyl cis-trans isomerase b precursor(cyclophilin b)	9e-37
P H10	Q9CR16	mus musculus (mouse). 4930564j03rik protein	2e-34
T G01	Q9XZZ5	lumbricus rubellus (humus earthworm). cyclophilin a	1e-65
J B09	Q9V3V2	drosophila melanogaster (fruit fly). fkbp13 protein	2e-16
E C08	P51121	xenopus laevis (african clawed frog). glutamine synthetase	2e-54
I E09	Q9I3S3	pseudomonas aeruginosa. agmatinase	1e-49
I D01	Q9VR81	drosophila melanogaster (fruit fly). cg17065 protein	4e-33
R D01	Q16775	homo sapiens (human). hydroxyacylglutathione hydrolase	1e-19
R E08	Q9Z2Z6	mus musculus (mouse). mcac protein	8e-18
E E03	O60547	homo sapiens (human). gdp-mannose 4,6 dehydratase	1e-31
I F02	P31153	homo sapiens (human). s-adenosylmethionine synthetase gamma form)	7e-49
O F08	Q9H0T9	homo sapiens (human). hypothetical 43.6 kda protein	5e-19
I F09	Q64424	myocastor coypus (coypu) (nutria). pancreatic lipase related protein 2 precursor	2e-14
H G02	O44391	strongylocentrotus purpuratus (purple sea urchin).ovoperoxidase	9e-09
R G09	Q9VGY2	drosophila melanogaster (fruit fly). cg6482 protein	2e-20
K H07	Q9C2G6	neurospora crassa. probable saccharopine reductase (lys3)	4e-25
R H07	Q9CSP7	mus musculus (mouse). 2700017m01rik protein (fragment)	5e-17
O H08	P78417	homo sapiens (human). glutathione-s-transferase homolog	9e-20
R A06	Q9R170	rattus norvegicus (rat). cytosolic branch chain aminotransferase bcatc (fragment)	4e-13
E B04	Q9Z2N9	rattus norvegicus (rat). phosphatidylcholine transfer protein	4e-37
H B12	Q9VPE2	drosophila melanogaster (fruit fly). cg6020 protein	2e-35
D C06	P07379	rattus norvegicus (rat). phosphoenolpyruvate carboxykinase, cytosolic	1e-43
J E12	Q9D1Q4	mus musculus (mouse). 1110001h19rik protein	7e-14
I G06	Q9ZFX7	pseudomonas alcaligenes. hypothetical 100.9 kda protein	3e-32
H H11	P50441	sus scrofa (pig). glycine amidinotransferase	2e-61
J F09	P53445	lampetra japonica (japanese lamprey). fructose-bisphosphate aldolase, muscle type	1e-65
A D10	P09367	rattus norvegicus (rat). l-serine dehydratase/l-threonine deaminase	1e-12
A F10	P11708	sus scrofa (pig). malate dehydrogenase, cytoplasmic	1e-47
B A08	Q18647	caenorhabditis elegans. similar to eosinophil peroxidase and myelo-peroxydase	6e-10
B C06	Q9H8H1	homo sapiens (human). cdna flj13639 fis, clone place1011219	3e-22
B D12	O94766	human. galactosylgalactosylxylosylprotein 3-beta glucuronosyltransferase 3	2e-34
B F07	P17244	cricetulus griseus (chinese hamster). glyceraldehyde 3-phosphate dehydrogenase	3e-44
B F11	P15651	rattus norvegicus (rat). acyl-coa dehydrogenase, short-chain specific precursor	1e-07
B H01	P29401	homo sapiens (human). transketolase	8e-52
C B02	Q9JK38	mus musculus (mouse). emeg32 (glucosamine-phosphate n acetyltransferase)	1e-27
E C06	Q9F2K1	streptomyces coelicolor. putative sugar hydrolase	6e-25
E H11	Q29305	sus scrofa (pig). disulfide isomerase-related protein (erp72) (fragment)	4e-11
F D12	P16460	mus musculus (mouse). argininosuccinate synthase	1e-79
G B12	P20258	pseudechis porphyriacus (red-bellied black snake). phospholipase a2	1e-14
G E05.	Q16698	homo sapiens (human). 2,4-dienoyl-coa reductase, mitochondrial precursor	6e-17
M B04	Q9JLZ3	mus musculus (mouse). au-binding enoyl-coa hydratase	1e-19
M G05	P11915	rattus norvegicus (rat). nonspecific lipid-transfer protein precursor	2e-63
N H08	Q9BVT4	homo sapiens. selenophosphate synthetase, human selenium donor protein	2e-24
P B03	Q9XYA1	limulus polyphemus (atlantic horseshoe crab). 5-aminolevulinate synthase	5e-33
P C07	Q9HNH6	halobacterium sp. (strain nrc-1). threonine dehydratase	4e-18
P E08	P97259	c alpha-1,3(6)-mannosylglycoprotein beta-1,6-n-acetyl-glucosaminyltransferase v	1e-20
P H06	O93662	methanosarcina barkeri. catalase	3e-38
S H12	Q9D635	mus musculus (mouse). 4733401p21rik protein	2e-47
T H07	Q03013	homo sapiens (human). glutathione s-transferase mu 4	4e-09

U B10	Q9CPU4	mus musculus (mouse). 2010306b17rik protein	1e-29
V B06	P45952	mus musculus (mouse). acyl-coa dehydrogenase, medium-chain specific precursor	6e-41
W G09	Q92626	homo sapiens (human). myeloblast kiaa0230 (fragment)	2e-29
Z D02	Q26710	trypanosoma brucei brucei. alternative oxidase precursor	3e-29
Z E07	Q9CMK9	pasteurella multocida. hypothetical protein pm0811	1e-08
Z G06	P23935	bos taurus (bovine). nadh-ubiquinone oxidoreductase 13 kda-b subunit	3e-30
Z H06	P10768	homo sapiens (human). esterase d	3e-13
W C01	O96299	drosophila melanogaster (fruit fly). sodh-2 protein	2e-50
D G09	Q9W2H1	drosophila melanogaster (fruit fly). cg10795 protein	2e-36
E E11	P97762	mus musculus (mouse). pap-1	4e-25
F E03	O13035	g proactivator polypeptide precursor	3e-18
M C09	Q9NQR4	homo sapiens (human). nit protein 2	4e-27
U F11	O15532	homo sapiens (human), and macaca mulatta (rhesus macaque). selenoprotein w	4e-07
N G02	Q9BSH5	homo sapiens (human). similar to riken cdna 2810435d12 gene	6e-27
F D07	Q9I2W7	pseudomonas aeruginosa. probable methyltransferase	8e-31
Q C02	O97427	drosophila melanogaster (fruit fly). eg eg0007.9 protein	4e-53
M C05	Q9BQ26	homo sapiens (human). unknown (carnitine palmitoyltransferaseii)	2e-19
R B06	Q9D445	mus musculus (mouse). 4933414e04rik protein	1e-22
I B10	O14874	homo sapiens (human). [3-methyl-2-oxobutanoate dehydrogenase (lipoamide)] kinase, mitochondrial precursor	1e-35
Thioredoxin/glutaredoxins:			
F A09	Q9H299	homo sapiens (human). p1725 (sh3bgrl3-like protein)	3e-10
P A10	P39811	escherichia coli. glutaredoxin 2 (grx2)	4e-36
E D10	Q9BH70	theileria parva. glutaredoxin-like protein	5e-11
A A03	Q9W1I7	drosophila melanogaster (fruit fly). cg5554 protein	3e-45
A B07	Q95108	bos taurus (bovine). mitochondrial thioredoxin precursor	2e-28
ATPase's:			
Y B02	Q9VP18	drosophila melanogaster (fruit fly). cg7625 protein	3e-18
P D07	Q9JL49	mus musculus (mouse). atp-binding cassette protein (fragment)	1e-18
M A10	O75964	homo sapiens (human). atp synthase g chain, mitochondrial	1e-14
C B12	P05630	bos taurus (bovine). atp synthase delta chain, mitochondrial precursor	3e-13
G F02	Q9VKM3	drosophila melanogaster (fruit fly). cg6105 protein	2e-16
S D09	Q24439	drosophila melanogaster (fruit fly). atp synthase oligomycin	1e-13
V G03	P05631	bos taurus (bovine). atp synthase gamma chain, mitochondrial precursor	8e-22
V A10	Q98TW9	ophiophagus hannah (king cobra) (naja hannah). atp synthetase beta-subunit	7e-37
T D10	Q9CQ23	mus musculus (mouse). atp synthase c chain (lipid-binding protein) (subunit c)	6e-07
O G04	P98197	mus musculus (mouse). potential phospholipid-transporting atpase ih	4e-37
Y C07	Q9UBP0	homo sapiens (human). spastin	1e-39
Cytochromes and related proteins:			
J D06	P15954	homo sapiens (human). cytochrome c oxidase polypeptide viic precursor	7e-08
R G06	P12787	mus musculus cytochrome c oxidase polypeptide va, mitochon. precursor	3e-34
B B01	Q9I8T9	thunnus obesus. cytochrome c oxidase subuniti v isoform 2 precursor	2e-14
C C11	Q9IAT5	meleagris gallopavo (common turkey). cytochrome c oxidase subunit via	4e-22
V B08	P81280	alligator mississippiensis (american alligator). cytochrome c	7e-48
W E05	Q9DF55	ophiophagus hannah (king cobra). cytochrome c oxidase copper chaperone	4e-16
E F12	Q9H3M8	homo sapiens (human). nadph-cytochrome p-450 reductase	1e-25
H D07	Q9D9F6	mus musculus (mouse). 1700082c19rik protein	1e-16
A H10	P23004	bos taurus (bovine). ubiquinol-cytochrome c reductase complex core protein 2 precursor (complex iii subunit ii)	4e-28

G. PROTEIN DEGRADATION AND PROCESSING

Proteasomes and its subunits:			
V D11	Q9V405	drosophila melanogaster (fruit fly). rpt3 protein	2e-38
X C12	Q99LL7	mus musculus (mouse). similar to proteasome 26s subunit, non-atpase, 3	3e-32
D D11	Q9UNM6	homo sapiens (human). 26s proteasome subunit p40.5, non-atpase, 13	3e-38
K C12	P48556	homo sapiens (human). 26s proteasome regulatory subunit s14 (p31)	7e-28
I D08	Q63347	rattus norvegicus (rat). 26s protease regulatory subunit 7 (mss1 protein)	4e-44
R G10	Q99460	homo sapiens (human). 26s proteasome regulatory subunit s1 (p112)	5e-10
C C07	Q9TZ67	caenorhabditis elegans. f40g9.1 protein	3e-11
E G06	P91729	geodia cydonium (sponge). lmp7-like protein	6e-51
F B04	Q9PUS1	brachydanio rerio (zebrafish) (zebra danio). proteasome subunit beta 7 (fragment)	3e-42
B E05	Q9BS70	homo sapiens (human). similar to proteasome subunit, alpha type, 3	5e-39
Proteases and peptidases:			
X D01	P29144	homo sapiens (human). tripeptidyl-peptidase ii	2e-72
A E06	Q9EPB1	rattus norvegicus (rat). dipeptidyl-peptidase ii precursor	5e-39
D A03	P53634	homo sapiens (human). dipeptidyl-peptidase i precursor	2e-57
O F01	Q9Z2W0	mus musculus (mouse). aspartyl aminopeptidase	1e-48

P G11	Q9BYZ4	homo sapiens (human). prtd-ny2	2e-11
E G09	Q9U554	tethya aurantia. silicatein beta	2e-09
A C07	Q9VP10	drosophila melanogaster (fruit fly). cg7169 protein	1e-13
N D01	Q9UF67	homo sapiens (human). hypothetical 35.5 kda protein (fragment)	8e-09
E B11	Q15005	homo sapiens (human). microsomal signal peptidase 25 kda subunit	9e-29
N A05	Q9ZDD3	rickettsia prowazekii. possible protease sohb (sohb)	2e-21
Ubiquitins:			
X F04	Q9WR78	bovine viral diarrhea virus-2 (bvdv-2). polyprotein (fragment)	5e-14
E A11	Q93116	acropora millepora. ub52	3e-34
J F05	Q93068	homo sapiens (human). ubiquitin-like protein smt3c	6e-22
Chaperone and heat shock proteins:			
H A08	P50503	rattus norvegicus (rat). hsc70-interacting protein	6e-10
O H10	P80317	mus musculus (mouse). t-complex protein 1, zeta subunit	4e-64
C E09	O42283	gallus gallus (chicken). 10 kda chaperonin	9e-29
F A06	Q9H1X3	homo sapiens (human). ba16l21.2.1 (novel dnaj domain)	6e-17
H D04	O73922	oreochromis mossambicus (mozambique tilapia). heat shock protein 70	2e-38
A H09	O73885	gallus gallus (chicken). heat shock cognate 70 kda protein	5e-54
C A12	P08110	gallus gallus (chicken). endoplasmin precursor (heat shock 108 kda protein)	3e-19

H. PROTEINS INVOLVED IN TRANSPORT

Ion transport and vitamin binding:			
P D01	P42577	lymnaea stagnalis (great pond snail). soma ferritin	4e-54
V G04	O61915	ixodes ricinus (castor bean tick). ferritin	8e-12
Q F12	P79849	trionyx sinensis (chinese softshell turtle). riboflavin binding protein precursor	7e-07
F E11	P04006	mus musculus (mouse).cysteine-rich intestinal protein	9e-22
1 A01	P39183	escherichia coli. amino acid antiporter (extreme acid sensitivity protein)	5e-71
K F02	O75555	homo sapiens (human). abc transporter moat-b isoform (fragment)	3e-23
E A10	Q9W3W0	drosophila melanogaster (fruit fly). cg4536 protein	3e-47
H G12	Q9NY16	homo sapiens (human). calcium channel alpha2-delta3 subunit	3e-12
I A03	O00244	homo sapiens (human). copper transport protein atox1	9e-12
X F03	O44930	aiptasia pallida. putative voltage-gated sodium channel	6e-09
Y C09	P55013	squalus acanthias.bumetanide-sensitive Na+-(K+)-Cl cotransporter1	4e-14
H E01	Q9JJH7	mus musculus (mouse). mlsn1-and trp-related protein 1 (mtr1)	4e-10
S F05	Q9UQ06	homo sapiens (human). bec1	4e-10
V E08	Q9QXA6	mus musculus (mouse). b(0,+)-type amino acid transporter 1	1e-28
T F09	Q09711	schizosaccharomyces pombe hypothetical ca-binding prot c18b11.04	4e-09
F H05	P81004	xenopus laevis (african clawed frog). voltage-dependent anion-selective channel protein 2 (outer mitochondrial membrane protein porin)	9e-45
Membrane trafficking and vesicle formation:			
E F11	Q9V754	drosophila melanogaster (fruit fly). cg10153 protein	2e-65
V F07	Q9NHE5	drosophila melanogaster (fruit fly). secretion calcium-dependent activator protein	2e-20
B H08	Q9CY18	mus musculus (mouse). 2510028h01rik protein	1e-17
B B12	Q9H1T7	homo sapiens (human). ba261p9.1.1 (syntaxin 16a / syntaxin 16b) (fragment)	1e-39
Z H07	Q9V3Y1	drosophila melanogaster (fruit fly). cg10130 protein	8e-14
1 C01	P38384	canis familiaris (dog). protein transport protein sec61 gamma subunit	1e-15
S H07	Q15436	homo sapiens (human). protein transport protein sec23 homolog isoform a	5e-30
V E03	Q99K49	mus musculus (mouse). sec23b (s. cerevisiae)	5e-21
A F02	Q9DCF9	mus musculus (mouse). 0610038p07rik protein	2e-56
R E04	Q62186	mus musculus (mouse). translocon-associated protein, delta subunit precursor	2e-28
E G01	Q9D7A6	mus musculus (mouse). 2310020d23rik protein	4e-24
S F03	Q9VZ29	drosophila melanogaster (fruit fly). cg2522 protein	1e-56
Z F04	P48444	homo sapiens (human). coatomer delta subunit	1e-33
I E12	O14579	homo sapiens (human). coatomer epsilon subunit	1e-44
P E10	O35643	mus musculus (mouse). beta-prime-adaptin protein	1e-23
H H07	P52303	rattus norvegicus (rat). beta-adaptin	6e-72
F F01	P11442	rattus norvegicus (rat). clathrin heavy chain	1e-29
O H03	Q9D2N9	mus musculus (mouse). 3830421m04rik protein	5e-30
1 D09	Q9UI05	homo sapiens (human). putative rab5-interacting protein (dj977b1.3.1)	7e-40
Nuclear import and export:			
S H03	Q9CT07	mus musculus (mouse). karyopherin (importin) alpha 3 (fragment)	8e-48
N B08	Q9UIJ5	homo sapiens (human). rec protein	3e-10
J H05	Q9IBE8	xenopus laevis (african clawed frog). ran gtp-binding protein	9e-09
B A09	P42558	gallus gallus (chicken). gtp-binding nuclear protein ran (tc4)	6e-81
H B08	Q9EQ30	mus musculus (mouse). ran binding protein 5 (fragment)	4e-10

I. SIGNALLING /REGULATORY MOLECULES

Extracellular and membrane-bound:			
W D03	Q25199	hydra attenuata (hydra) (hydra vulgaris). tyrosine kinase receptor	6e-15

G A12	Q13332	homo sapiens (human). protein-tyrosine phosphatase, receptor-type, s precursor	5e-10
U H04	Q13221	homo sapiens (human). cysteine-rich fibroblast growth factor receptor	1e-11
Y G01	P48800	gallus gallus (chicken). heparin-binding growth factor 2 precursor	1e-20
Q B03	O57409	brachydanio rerio (zebrafish) (zebra danio). deltab	2e-07
F F08	Q98UF7	fugu rubripes (japaenese pufferfish) pecanex	4e-17
S B11	O13149	fugu rubripes (japanese pufferfish) (takifugu rubripes). notch 2 (fragment)	3e-07
P F07	P97806	mus musculus (mouse). oncoprotein induced transcript 3 (ef-9) (fragment)	4e-14
U B02	P35440	gallus gallus (chicken). thrombospondin 1 precursor	1e-17
H E10	P35442	homo sapiens (human). thrombospondin 2 precursor	6e-10
K B03	O62542	geodia cydonium (sponge).endothelial-monocyte-activating polypeptide related prot.	3e-33
1 B08	Q9D0I6	mus musculus (mouse). 2610014f08rik protein	3e-17
R H03	Q9V8M4	drosophila melanogaster (fruit fly). cg15078 protein	6e-16
D B04	O75629	homo sapiens (human). cellular repressor of e1a-stimulated genes creg	1e-32
Z E04	P06734	homo sapiens (human). low affinity immunoglobulin epsilon fc receptor	7e-07
X B08	P30205	bos taurus (bovine). antigen wc1.1	2e-22
Y E05	Q05049	xenopus laevis (african clawed frog). integumentary mucin c.1 (fim-c.1) (fragment)	4e-18
Q B05	Q9N1T0	ornithorhynchus anatinus (duckbill platypus) mannose 6-phosphate/insulin-like growth factor 2 receptor	3e-10
Intracellular:			
I B09	Q9NL95	hemicentrotus pulcherrimus (sea urchin). ck2 alpha subunit	3e-23
E G02	O57468	xenopus laevis (african clawed frog). 14-3-3 protein epsilon	2e-59
C F01	Q9V7F0	drosophila melanogaster (fruit fly). cg8242 protein	1e-29
C B05	Q9VQG5	drosophila melanogaster (fruit fly). cg2862 protein	9e-37
I E08	O15992	anthopleura japonicus (sea anemone). arginine kinase	5e-35
B C05	Q9IBG6	cyprinus carpio (common carp). glia maturation factor beta	5e-43
O H11	Q9KHA3	nostoc punctiforme. putative hexuronic acid kinase hrmk	7e-17
B F07	Q9VZV9	drosophila melanogaster (fruit fly). cg1271 protein	3e-21
D H04	Q9H3S4	homo sapiens (human). thiamin pyrophosphokinase	2e-25
X B06	Q9Y6K8	homo sapiens (human). adenylate kinase isoenzyme 5	6e-15
Y F10	P13686	homo sapiens (human). tartrate-resistant acid phosphatase type 5 precursor	1e-10
E E12	Q9D892	mus musculus (mouse). 2010016i08rik protein	2e-21
C E06	Q9Z1T6	mus musculus (mouse). fyve finger-containing phosphoinositide kinase	3e-42
E E01	O76039	homo sapiens (human). serine/threonine-protein kinase 9	2e-35
X G09	Q9VYI2	drosophila melanogaster (fruit fly). cg4400 protein	5e-08
W A08	Q9VMF3	drosophila melanogaster (fruit fly). cg9491 protein	3e-34
C H04	O35413	rattus norvegicus (rat). sh3-containing protein p4015	1e-10
P E05	Q9BW87	homo sapiens (human). similar to vaccinia virus hindiii k4l orf	1e-13
F D01	P04080	homo sapiens (human).cystatin b(liver thiol proteinase inhibitor)(cpi-b) (stefin b)	4e-19
E D05	P41044	drosophila melanogaster (fruit fly). calbindin-32	4e-20
I C12	Q9DCT5	mus musculus (mouse). stromal cell derived factor 2	1e-30
U E06	Q01066	rattus norvegicus (rat). calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1b	4e-10
Calmodulins:			
J G02	O96081	halocynthia roretzi (sea squirt). calmodulin b	3e-76
Q F03	O96949	geodia cydonium (sponge). calmodulin	4e-25
V A05	P02596	metridium senile (sea anemone), and renilla reniformis (sea pansy). calmodulin	9e-28
F H02	Q9D6G4	mus musculus (mouse). adult male hippocampus cdna, riken full-length enriched library, clone 2900055d23	3e-21

J. OTHERS

Fluourescent:			
Q H12	Q9U6Y8	discosoma sp. flourescent protein fp583	1e-33
B B04	Q9U6Y5	zoanthus sp. flourescent protein fp506	8e-82
M F09	Q9U6Y7	discosoma striata. flourescent protein fp483	6e-21
Proteins involved in cell-cycle regulation:			
K A09	Q9CR47	mus musculus (mouse). 5730427n09rik protein	1e-40
I F12	Q99547	homo sapiens (human). m-phase phosphoprotein 6	2e-18
A G09	Q9AVH6	pisum sativum (garden pea). putative senescence-associated protein	2e-37
S C03	Q9H8V3	homo sapiens (human). flj13205 fis,clone nt2rp3004534,similar to mouse ect2	7e-26
C C02	P79741	brachydanio rerio (zebrafish) (zebra danio). pes	6e-07
Miscellaneous:			
H E08	Q9CPZ2	mus musculus (mouse). 2310008m10rik protein	3e-16
R C11	Q9C0B0	homo sapiens (human). kiaa1753 protein (fragment)	1e-55
V G07	Q9JIX0	mus musculus (mouse). e(y)2 homolog	1e-24
X B11	Q9PVQ2	xenopus laevis (african clawed frog). loocyte membrane protein	3e-24

M F03	Q9U852	diadema savignyi. reverse transcriptase (fragment)	3e-12
I E02	Q9I7H1	drosophila melanogaster (fruit fly). cg18289 protein	3e-15
Z C01	Q9ZWT4	ipomoea purpurea (common morning-glory). transposase	1e-07
G F10	O94980	homo sapiens (human). kiaa0906 protein (fragment)	3e-17
Unknown:			
S F06	Q61856	mus musculus (mouse). orf	3e-21
W A11	Q9UQN0	homo sapiens (human). cgi-57 protein	5e-42
B C11	Q19269	caenorhabditis elegans. f09e8.6 protein	1e-06
K G10	Q13454	homo sapiens (human). n33 protein	3e-26
E B03	Q9H5V4	homo sapiens (human). cdna flj22986 fis, clone kat11742	1e-37
S G04	Q9ES77	mus musculus (mouse). polydom protein precursor	7e-12
R A04	O16003	hydra attenuata (hydra). collagen-like protein (fragment)	9e-07
1 A08	O14730	homo sapiens (human). sudd protein	7e-31
J B02	O95214	homo sapiens (human). brain my047 protein	1e-35
Z H12	Q9H3I2	homo sapiens (human). brain my044 protein (tetratricopeptide repeat domain 4)	1e-33
X G01	Q9D958	mus musculus (mouse). 1810004f21rik protein	2e-16
K F09	Q9Y244	homo sapiens (human). hspc036 protein (hypothetical 15.8 kda protein)	8e-31
P F05	Q9BXW1	homo sapiens (human). pnas-125	5e-07
V E04	Q9CSQ7	mus musculus (mouse). testis expressed gene 9 (fragment)	1e-12
H E11	Q9D1L0	mus musculus (mouse). ethanol induced 6	5e-25
B C12	Q9D172	mus musculus (mouse).dna segment,chr10,johns hopkins university 81expressed	7e-32
B E02	Q9BXF3	homo sapiens (human). cecr2 protein	2e-42
B F05	Q13438	homo sapiens (human). protein os-9 precursor	3e-16
B H03	Q9NAE2	caenorhabditis elegans. y51h4a.7 protein	6e-15
C E08	Q15055	homo sapiens (human). kiaa0033 protein (fragment)	1e-39
E D02	Q9VQA7	drosophila melanogaster (fruit fly). cg16995 protein	4e-14
P B11	Q9W4M9	drosophila melanogaster (fruit fly). cg6133 protein	2e-47
F H12	Q9NTW8	homo sapiens (human). dj794i6.1.1 (fragment)	1e-07
T E03	Q9N5Y2	caenorhabditis elegans. c15f1.6 protein	5e-19
T F12	Q9PTJ3	brachydanio rerio (zebrafish) (zebra danio). lmo2 protein	3e-09
U B06	Q9BX38	homo sapiens (human). ba528a10.3.1 (fragment)	1e-25
U D01	O80889	arabidopsis thaliana (mouse-ear cress). t26b15.8 protein	7e-28
U F02	Q9Y359	homo sapiens (human). cgi-43 protein	5e-44
V A06	O75050	homo sapiens (human). kiaa0462 protein (fragment)	2e-39
V B10	Q10206	schizosaccharomyces pombe (fission yeast). hypothetical 55.6 kda prot c17d1.02	4e-30
V H08	Q9UPR3	homo sapiens (human). kiaa1089 protein (fragment)	8e-09
W H02	Q9VAY7	drosophila melanogaster (fruit fly). cg12259 protein	2e-25
W H07	Q9BXB6	homo sapiens (human). nyd-sp14	2e-14
X C06	Q9H0P5	homo sapiens (human). hypothetical 90.1 kda protein	4e-16
X D12	P14588	plasmodium falciparum hypothetical protein 3' to asp-rich and his-rich proteins	2e-18
Y D08	P30042	homo sapiens (human). es1 protein homolog precursor	5e-25
K C03	Q9BZE7	homo sapiens (human). evg1	2e-08
C C12	Q9D2J7	mus musculus (mouse). adult male testis cdna, riken full-length enriched lib, clone 4921532c19	9e-48
Hypothetical:			
1 A02	Q9VXH4	drosophila melanogaster (fruit fly). cg9921 protein	3e-14
E A08	O70349	mus musculus (mouse). hypothetical 136.7 kda protein	4e-21
O A08	Q9VBN7	drosophila melanogaster (fruit fly). cg4743 protein	4e-08
E B01	Q9BS18	homo sapiens (human). unknown (protein for mgc 12537)	2e-13
H B01	Q9DB98	mus musculus (mouse). 1500001k17rik protein	1e-06
I B01	O76897	drosophila melanogaster (fruit fly). eg 73d1.1 protein	1e-17
E B07	Q9CSK8	mus musculus (mouse). 1810019e15rik protein (fragment)	5e-09
1 E09	P03845	escherichia coli. hypothetical protein 1	3e-11
1 B03	Q9DAP6	mus musculus (mouse). 1700003m02rik protein	5e-24
1 F06	Q9CQU0	mus musculus (mouse). 0610040b21rik protein	2e-34
Q D02	Q9HCV1	homo sapiens (human). y214h10.2 (kiaa0187 protein)	1e-18
D E02	Q9VLV6	drosophila melanogaster (fruit fly). cg7102 protein	5e-09
1 C01	Q9VPB4	drosophila melanogaster (fruit fly). cg3698 protein	6e-15
1 H04	Q9CQ86	mus musculus (mouse). 1810046j19rik protein	4e-10
O G09	O75140	homo sapiens (human). kiaa0645 protein (wuGSc h_dj403e2.1 protein)	8e-12
E H01	Q9D472	mus musculus (mouse). 4933409d10rik protein	7e-23
D H07	Q9BTC7	homo sapiens (human). unknown (protein for image3543931) (fragment)	2e-18
O H02	Q9BQM3	homo sapiens (human). dj842g6.1.1 (novel protein) (fragment)	1e-08
H H03	Q9D7L6	mus musculus (mouse). 2310003l22rik protein (riken cdna 2310003l22 gene)	1e-30
K B06	Q9Y3E3	homo sapiens (human). cgi-145 protein	1e-07
R C06	Q9R095	rattus norvegicus (rat). kpl2	1e-12
A C05	Q9CX74	mus musculus (mouse). 4122402o22rik protein	5e-20
A F12	Q9CQ22	mus musculus (mouse). 2400001e08rik protein	4e-23

A G07	Q14691	homo sapiens (human). hypothetical protein kiaa0186	1e-16
B A02	Q9BWJ5	homo sapiens (human). similar to cg11985 gene product	1e-28
B B11	Q9CSH5	mus musculus (mouse). 1300013g12rik protein (fragment)	1e-27
B D05	Q9T0A1	arabidopsis thaliana (mouse-ear cress). hypothetical 49.7 kda protein	5e-29
B E01	Q9CYV5	mus musculus (mouse). 2810439k08rik protein	7e-17
B H09	Q9D552	mus musculus (mouse). 4930513f16rik protein	1e-11
C B08	Q9BX05	homo sapiens (human). ba384d7.1.1 (dkfzp434p106 protein, isoform 1)	6e-30
C H09	Q93200	caenorhabditis elegans. w04g3.8 protein	1e-07
E A09	Q9VUC2	drosophila melanogaster (fruit fly). cg6451 protein	2e-12
E B10	Q9BV87	homo sapiens (human). hypothetical 45.4 kda protein	9e-40
E D07	Q9SIZ7	arabidopsis thaliana (mouse-ear cress). at2g22020 protein	9e-20
F B06	Q9P223	homo sapiens (human). kiaa1505 protein (fragment)	3e-22
F A07	Q9VHI4	drosophila melanogaster (fruit fly). cg11985 protein	1e-32
F F07	Q9D4V3	mus musculus (mouse). 4930554c01rik protein	2e-11
G A10	Q9H8Z5	homo sapiens (human). cdna flj13120 fis, clone nt2rp3002682	9e-13
I A04	O14037	schizosaccharomyces pombe (fission yeast). hypothetical 14.6 kda protein c2c4.04c	4e-10
M C04	Q9UF22	homo sapiens (human). cn113a11.1 (kiaa0645) (fragment)	2e-12
M H09	Q9CR92	mus musculus (mouse). 4921513e08rik protein	6e-32
N A04	Q9LYH2	arabidopsis thaliana (mouse-ear cress). hypothetical 56.8 kda protein	2e-07
N C08	Q9P207	homo sapiens (human). kiaa1521 protein (fragment)	1e-51
N D10	Q9VNQ3	drosophila melanogaster (fruit fly). cg11137 protein	3e-10
N H11	Q9Y439	homo sapiens (human). hypothetical 55.4 kda protein	3e-25
S D05	Q9HC36	homo sapiens (human). hypothetical 31.0 kda protein	1e-15
S E08	Q9V4K7	drosophila melanogaster (fruit fly). cg11166 protein	3e-17
U C02	Q9CQK5	mus musculus (mouse). 0610010i12rik protein	5e-21
U E07	O75165	homo sapiens (human). kiaa0678 protein (fragment)	7e-22
U H02	Q9CSU0	mus musculus (mouse). 2610304g08rik protein (fragment)	4e-27
V D12	Q9DB60	mus musculus (mouse). 2810405k02rik protein	4e-32
V G01	Q9D296	mus musculus (mouse). 5830411g16rik protein	2e-25
W D02	Q9UA16	strongyloides stercoralis. l3nieag.01 (fragment)	6e-13
X C08	Q9H6Q8	homo sapiens (human). cdna flj21977 fis, clone hep05976	1e-09
Y B01	Q9VXR1	drosophila melanogaster (fruit fly). cg12379 protein	1e-21
Y C04	Q19527	caenorhabditis elegans. f17c8.3 protein	5e-24
Y C06	Q9H5P8	homo sapiens (human). cdna flj23189 fis, clone lng12061	3e-17
Y G05	P90910	caenorhabditis elegans. k07a1.3 protein	6e-09
Z G03	O62198	caenorhabditis elegans. f32a11.1 protein	1e-09

Appendix A.5 Genomic sequence of the *Drosophila* CG7056 gene

The CG7056 locus incorporates two exons, shown in bold capitals

agtggagtggagtgccatcgaggcagtggagcgcctaaccaaccaccactcgcttatcaatcgatggccggtagc
tcggcactatctcccatgcgcggcatatctacataccgctctattattgttaatattgatatctacccattagca
caacattataggtacaatctattatattctgtgaaatcagatttggtacaatttatagcgcccatgggctctgcg
acggctcctacagctggtggccattagttcatcgcctgcggccgacgcagtcgcttcgatcgctccggacccaat
caatcggtatctacagttcgttactaccgctctcagaatccaccaatatggatatcgccaccaagtccagcaagg
ccgccttctccattgagaatatcttggagcagaagagctcctcgcatcgatctcaatcgcgcagaggctcctctc
aaagtcccgtggcggtaacagcaggcatagctactccacacgcactggcc**ATGCCCAAGGCATCCTTGGCC**
ACTGGATCATCCTCCGCGGCGCCTACACCCTCTCCATCGTCCGCCACCAACATCTACGATCTAT
CCCGCGAGGCAGCTGCCGCCCAGTATGCCATGAAGAGTATGGACTCCAGTGCTGTCCTGGCTCC
CACTTCGCTGCGCTTCAATCCCATCTACCCGGATCCCGCCTCGCTCTTCTACCAGCAGGTGCTG
CAGCTCCAGAAGAATCCCTCGCTCTTCATGCCCCACTTCCAGGCGGCCGCTGTTGCCGCCGCAG
CCGCTGTTCAGCCAACAGCCTATTGTGATCAATACAGTCCATTTACCATGGACTGCGAAGGTGAG
TACTTCAGAATACCTACTTAGTGGGGTTAACATGTATTTTAGTTGGGTGTTGAATGTCCTAAAAGATAAAGATTC
ATTTTCAAGATAATTGCTTAATTTTTGTGGTAATTTATGATAAATCATTCACTAAAATAAGGGTGCATGAGGAGA
AGACATTACAAATTTATTATAAAGTTATGGTATTTCCACAGAGCTATAGAAAAGGTATAGAAATGAAATATTAAA
TTATATGCAATCTAATCTCAGCAAATTCAAAAGATTGTAAAGCAGTTATAAGGCCAAGAAGTATAGTATATTTAT
TCGACATTATGAATGAAGACAATACAGCATTTCTTTATTAAAAAAAATTCTAGAATAAGATCAAAATTATTCAAT
TATTCGAGCTGTATACATACATGCAAACGCAGTTGAAACCCTTAGGTCAAGTTATTGACAAGCTCATAAAATAAT
TGACTAATATATTAATTAATGCGGTTTTGAGAAATCTATTAGCCCAATGATGTCTATCAACAATTTAGTCGTTTG
ACCTCATTAGCATATTTCATCATACTATCAAATGACCGGAAATGAGTCACCTGAATATGAATGAATAAGCGCTAC
AATTATCAATTAAACACTAAAATGCAGATAAAGGCAGAAGCATTCTCAGAGTATGATATGTTAAATCCCTATTCT
TAAATTCTTATTCATGAAGCTTGGAACCATAGTTTCTATTAGACTTTCTCACAGTTTTTAGCAAAGACTAGGATT
GATTCGACACTACCAAGTTCAAGTCATTGAACCTACCCATGCTGACACCTTAATCTCGCTCATCAAGTTGACACA
TGGAAATTGATCGCCATGGGAAAGGTGTTGCCTGAAATCCCCATTTAAATACCAAATATCTAAGCTAGCCTAAGT
CCGAAGTGCCCGGACCACCGAATATTACAATGGGGATCCCACACAAAATGGGCCATAAACATGATTAATGTGTGT
GTGTTGTGGTGCAGTTGGGTATTTGTCAAAGTCAAGTGTTGGCGCCAATGCCACTCGAAATCACATTTTCCCCGC
CCCAAAAACTGCATTTCCCGGTTGGTTTTCTGGGCGGTGGTGTAACGGTAGATTAGCCCTAGCAATGCTGATAAA
GTCAAAGTCCCGAATTCCATCCCCTCGCAGTACAAATGCCCCATTCAGATCGGAAAATATTTCGGTGGCTTATCA
GAAGTGACAAATATTTGAAATTCAGGTGAGAGACAGCAGAAACTATATCTATGGATATTTAAGTGACAGTTAGTT
TTCATTGAAAGGCATTGAGCTGTTGATTTGGGAAAATCGAGCTGCTGCGAATTAGACGGACTGTGAGAAATCATA
CTCATTTAAACCTTTGGATAACACACAATTATCACAGGCAGTATCTTTCGATATAATTGTAGTGAAATAGGAAGG
AATCAATACAACATTATAGTGGAATGGAATATCTAAGTGCACCTTTCTGTTCCCAATATGCCCGTCTTGAGGATA
TATCTATGATGAGTCCAACTGAGTTAGGATCATACTACGCCTCAAAAAAAAAAAAGGAGTATTGATTAGCTTGCT
CTATGCCATACTGATCACACAATGCAGATAGGAGTATCGTCGTGTATAGGGCACTTCAACTGTACTGCTATCCCG
ATTTCCGATTTGTTTGGCTTATCTAAACAAAAATGCTATTAGAGGCGATTAATAGGCGCTGTCCATCATCTTATA
GCCACGATTGCCATTTGATTACGATAAGCCCGTCCGTTCCCGACACTCGTTCCACCTGGACAAGGGGCAAATGGT
TAGGGGCGGAGGCGGGTGTGGTGGTGTTGGCGACAAGTACCTGATATGATATTATTCCTTCAATCGGACACGCAC
ATTATAAAATAATAACAAATCCAATTTTCAGCGATTAAATGCCGCTGGGAAGAAGATAAGGGTGCTCTGCGGCCT
TATCACTCCCGAATACGCCACTTAATAAAAATAAACATGAATAAGTCTATTTGTAACGATTTAAATTATGGCATC
TGCAATCGGATGTTATTTATTTAATGGCATTTGGCCATTTCCAAGGGTTCTTGGCCAGATCGCCTCACCAATGAG
GTAACTGGGTCCCCAAATATCAGCTTAGTTCAGCGATGGAGCACACTCCAAATATGCGCCAAGTGCGAATTTCCG
TTACCTTAACTAAAAAAAAAAAAAAGAGGCCACCTTCCTTATCACATAGGTGTATGTACATAGGTGCAGATAGGTT
CGCAATCGGATTGCCGTTCGGGTGCCAATTGATTTCGCCCAATTGCTACGAAATCTCGGCGTACAGTCGATCCAT
TTCGCAATGGCCACCATTCGCAAGTCGTCATCCAGTTACAAGTGGGCGTTGACATATTCTCCGATACTGATACAA
CACCCATACACTTGTACCACTCGCAATTGTTGTGTCCGATTACTTTGATAAGAGCGGGTGGGCGGAGATAGGGAA
AACCGGGCCAATGGAGCCACCAATCACAATCCAATGATTGCGCCACTTTATATCGGGCAGATGGTTCAAAATCGT
TATTCAAAGAAATCGATTGTCCTCGACTTTTTCATGGCAGCTTTTGATAAACTCTGAAGTGGGTCTATAATGTCA
TTAGTTCTTTAATTAAGATCGTATTAACAATGAAATATTTAGTTTAAATATTTCTCATCTGATATAGGAAACCCT
TGTTCACAAAGTGAAGTGTTAACAATAAGTTAGTTCTGTAACCTCTTCAGCAATTGTAATCTAAAACAAACTGTC
TTCCTTTTTAATCCATAG**GATTCCCCAATCCTGCCTCCGCTGCAGCCGCCCTCTACTGCAATGCCTA**
TCCGGCGGCCAGCTTCTACATGTCGAACTTCGGCGTGAAGCGAAAAGGCGGCCAGATCAGATTC
ACCTCCCAGCAGACCAAAAATCTCGAGGCTCGCTTCGCCAGCTCCAAGTACCTGTCGCCGGAGG
AACGGCGTCATCTGGCCCTCCAACTCAAGCTGACCGACCGCCAGGTGAAGACCTGGTTTCAGAA
CCGGCGCGCCAAGTGGCGACGAGCCAATCTCAGCAAGCGCAGTGCATCCGCCCAAGGACCCATA
GCAGGCGCTGCCGTGGGATCACCCTCCAGCGCCTCGAGCAGCAGTGTTCCCGTGTTGAATCTTG
GCAGCGGAAGCAGGTGTGGCCAACAAAGCGACGAGGAGGATCGGATGTATCTGTCCGAGGACGA
TGAGGACGACGACGAAGACGAAGGCGAGGCGGATGAGACGCCCAAGTGAagctagaggatggcccat

tgactcataagttagtttagcttgtagttaggataaataaatttattttattagaactatagttatgcgagttag
taagtactagatgggtggtctcctagtggaactggtagctgtccaagcaggcgcaattgggcaaggtgcatcttg
gcgccagattggagaggtgacccatggtcaccggaaaggatagcggcacaaaccacttggggtacgcacaaatgg
ccgctgccgccgccgctgatgcactgcctccaaagaccggcggctttagctccaccaccgatttgctctggtact
gaaaggtaagattttgtatgtacttaaataaatatttcattaggtacttcctacttccatggaccacgcaaatcg
tgacttgattgaattctggaagcggaaatacagaccgttgaacttaagtaaaaacacaaaacacctgttacagat
ttgctgtatactatagaatataatactataga

References

Adachi, J. and Hasegawa, M. (1996). "MOLPHY version 2.3 programs for molecular phylogenetics bases on maximum likelihood." Institute of Statistical Mathematics, Tokyo.

Adoutte, A., Balavoine, G., Lartillot, N. and de Rosa, R. (1999). "Animal Evolution; the end of intermediate taxa?" Trends in Genetics **15**(3): 104-108.

Ajioka, J. W., Boothroyd, J.C., Brunk, B.P., Hehl, A., Hillier, L., Manger, I.D., Marra, M., Overton, G.C., Roos, D.S., Wan, K.L., Waterston, R. and Sibley, L.D. (1998). "Gene discovery by EST sequencing in *Toxoplasma gondii* reveals sequences restricted to the Apicomplexa." Genome Res. **8**(1): 18-28.

Altschul,T. M., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucl. Acids. Res. **25**: 3389-3402.

Angerer, L. M., Oleksyn, D. W., Logan, C. Y., McClay, D. R., Dale, L. and Angerer, R. C. (2000). "A BMP pathway regulates cell fate allocation along the sea urchin animal-vegetal embryonic axis." Development **127**: 1105-1114.

Arendt, D. and Nubler-Jung, K.(1994). "Inversion of dorsoventral axis?" Nature **371**(6492): 26.

Arendt, D. and Nubler-Jung, K. (1999). "Comparison of early nerve cord development in insects and vertebrates." Development **126**(11): 2309-25.

Arora, K. and Nusslein-Volhard, C.(1992). "Altered mitotic domains reveal fate map changes in *Drosophila* embryos mutant for zygotic dorsoventral patterning genes." Development **114**(1003-1024).

Attisano, L., Carcamo, J., Ventura, F., Weis, F. M., Massagué, J. and Wrana, J. L. (1993). "Identification of human activin and TGFβ type I receptors that form heteromeric kinase complexes with type II receptors." Cell **75**: 671-680.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1995). Current Protocols in Molecular Biology.

Azpiazu, N. and Frasch, M. (1993). "*tinman* and *bagpipe*: two homeo box genes that determine cell fates in the dorsal mesoderm of *Drosophila*." Genes Dev. **7**: 1325-1340.

Baker, J. and Harland, R. M. (1996). "A novel mesoderm inducer, Madr2, functions in the activin signal transduction pathway." Genes Dev. **10**: 1880-1889.

Baker, N. E. (2000). "Notch signaling in the nervous system. Pieces still missing from the puzzle." Bioessays **22**(3): 264-273.

Ball, E. E., Hayward, D., Reece-Hoyes, J., Hislop, N., Samuel, G., Saint, R., Harrison, P. and Miller, D. J. (in press). "Coral Development: from Classical Embryology to Molecular Control." IJDB.

Beddington, R. S. and Robertson, E. J. (1999). "Axis development and early asymmetry in mammals." Cell **96**: 195-209.

Bedford, F. K., Ashworth, A., Enver, T. and Wiedemann, L.M. (1993). "HEX: a novel homeobox gene expressed during haematopoiesis and conserved between mouse and human." Nucleic Acids Res. **21**(5): 1245-1249.

Berghammer, H., Hayward, D., Harrison, P. and Miller, D.J. (1996). "Nucleotide sequence of *ub52* from the cnidarian *Acropora millepora* reveals high evolutionary conservation." Gene. **178**(1-2): 195-197.

Bienz, M. (1994). "Homeotic genes and positional signalling in the *Drosophila* visceral mesoderm." Trends. Genet. **10**: 22-26.

Bilder, D. and Scott, M. P. (1995). "Genomic regions required for morphogenesis of the *Drosophila* embryonic midgut." Genetics **141**: 1087-1100.

Bodmer, R., Jan, L. Y.and Jan, Y. N. (1990). "A new homeobox-containing gene, *msh-2*, is transiently expressed early during mesoderm formation of *Drosophila*." Development **110**: 661-669.

Bodmer, R. (1993). "The gene *tinman* is required for specification of the heart and visceral muscles in *Drosophila*." Development **118**: 719-729.

Bogue, C. W., Ganea, G. R., Strum, E., Ianucci, R. and Jacobs, H. C. (2000). "Hex expression suggests a role in the development and function of organs derived from foregut endoderm." Dev. Dyn. **219**(1): 84-89.

Brand, A. H. and Perrimon, N. (1993). "Targeted gene expression as a means of altering cell fates and generating dominant phenotypes." Development **118**(2): 401-15.

Brickman, J. M., Jones, C. M., Clements, M., Smith, J. C. and Beddington, R. S. P. (2000). "Hex is a transcriptional repressor that contributes to anterior identity and suppresses Spemann organiser function." Development **127**: 2303-2315.

Bridge, D., Cunningham, C. W., Schierwater, B., DeSalle, R. and Buss, L. W. (1992). "Class level relationships in the phylum Cnidaria." PNAS USA **89**: 8750-8753.

Bridge, D., Cunningham, C. W., DeSalle, R. and Buss, L. W. (1995). "Class-level relationships in the phylum Cnidaria: Molecular and morphological evidence." Mol. Biol. Evol. **12**(4): 679-689.

Brower, D. L., Brower, S. M., Hayward, D. C. and Ball, E. E. (1997). "Molecular evolution of integrins: genes encoding integrin beta subunits from a coral and a sponge." PNAS USA **94**: 9182-9187.

Brummel, T. J., Twombly, V., Marques, G., Wrana, J. L., Newfeld, S. J., Atissano, L., Massagué, J., O'Connor, M. B. and Gelbart, W. M. (1994). "Characterization and realationship of DPP receptors encoded by the *saxophone* and *thickvein* genes in *Drosophila*." Cell **78**(251-261).

Bryant, P. J. (1988). "Localized cell death caused by mutations in a *Drosophila* gene coding for a transforming growth factor-beta homolog." Dev. Biol. **128**: 386-395.

Campos-Ortega, J. A. and Vassin, H. (1985). The Embryonic development of *Drosophila melanogaster*, Springer-Verlag, Berlin.

Capovilla, M., Brandt, M. and Botas, J. (1994). "Direct regulation of *decapentaplegic* by *Ultrabithorax* and its role in *Drosophila* midgut morphogenesis." Cell **76**(461-475).

Chen, R.-H. and Derynck, R. (1994). "Homomeric Interactions between Type II Transforming Growth Factor-β Receptors." J. Biol. Chem. **269**(36): 22868-22874.

Chen, R.-H., Moses, H. L., Maruoka, E. M., Derynck, R. and Kawabata, M. (1995). "Phosphorylation-dependent Interaction of the cytoplasmic Domains of the Type I and Type II Transforming Growth Factor-β Receptors." J. Biol. Chem **270**(20): 12235-12241.

Chen, Y., Rubock, M. J. and Whitman, M. (1996). "A transcriptional partner for MAD proteins in TGF-β signaling." Nature **3834**(691-696).

Chen, Y. G., Hata, A., Lo, R. S., Wotton, D., Shi, Y., Pavletich, N. and Massague, J. (1998). "Determinants of specificity in TGF-beta signal transduction." Genes Dev. **12**(14): 2144-2152.

Chen, Y. G. and Massagué, J. (1999). "Smad1 recognition and activation by the ALK1 group of transforming growth factor-beta family receptors." J. Biol. Chem. **274**(6): 3672-3677.

Christian, J. L. and Nakayam, T. (1999). "Can't get no SMADisfaction: Smad proteins as positive and negative regulators of TGF-β family signals." Bioessays **21**(5): 382-390.

Clement, J. H., Fettes, P., Knochel, S., Lef, J. and Knochel, W. (1995). "Bone morphogenetic protein 2 in the early development of *Xenopus laevis*." Mech. Dev. **52**: 357-370.

Crompton, M. R., Bartlett, T. J., MacGregor, A. D., Manfioletti, G., Buratti, E., Giancotti, V. and Goodwin, G. H. (1992). "Identification of a novel vertebrate homeobox gene expressed in haematopoietic cells." Nucl. Acids. Res. **20**: 5661-5667.

Dale, L. and Slack, J. M. W. (1987a). "Regional specification within the mesoderm of early embryos of *Xenopus laevis*." Development **100**: 279-295.

Dale, L. and Slack, J. M. W. (1987b). "Fate map for the 32-cell stace of *Xenopus laevis*." Development **99**: 527-551.

Dale, L. and Wardle, F.C. (1999). "A gradient of BMP activity specifies dorsal-ventral fates in early *Xenopus* embryos." Semin. Cell Dev. Biol. **10**(3): 319-326.

de Caestecker, M. P., Hemmati, P., Larisch-Bloch, S., Ajmera, R., Roberts, A. B. and Lechleider, R. J. (1997). "Characterization of Functional Domains within Smad4/DPC4." J. Biol. Chem. **272**(21): 13690-13696.

de Celis, J. F. (1997). "Expression and function of *decapentaplegic* and *thickveins* during the differentiation of the veins in the *Drosophila* wing." Development **124**(1007-1018).

DeRobertis, E. M. and Sasai, Y. (1996). "A common plan for dorsoventral patterning in Bilateria." Nature **380**(6569): 37-40.

Dosch, R., Gawantka, V., Delius, H., Blumenstock, C. and Niehrs, C. (1997). "Bmp-4 acts as a morphogen in dorsoventral mesoderm patterning in *Xenopus*." Development **124**: 2325-2334.

Ebner, R., Chen, R-H., Lawler, S., Zioncheck, T. and Derynck, R. (1993). "Determination of type I receptor specificity by type II receptors for TGFβ and activin." Science **262**: 900-902.

Eppert, K., Scherer, S. W., Ozcelik, H., Pirone, R., Hoodless, P., Kim, L-C., Tsui, L. C., Bapat, B., Gallinger, S., Andrulis, I. L., Thomsen, G. H., Wrana, J. L. and Attisano, L. (1996). "MADR2 maps to 18q21 and encodes a TGF beta-regulated MAD-related protein that is functionally mutated in colorectal carcinoma." Cell **86**(4): 543-552.

Fainsod, A., Steinbeisser, H. and de Robertis, E. M. (1994). "On the function of Bmp-4 in patterning the marginal zone of the *Xenopus* embryo." EMBO J **13**: 5015-5025.

Feng, J. Q., Chen, D,M., Cooney, A. J., Tsai, M-J., Harris, M. A., Tsai, S. Y., Feng, M. F., Mundy, G. R. and Harris, S. E. (1995). "The mouse Bone Morphogenetic Protein-4 gene." J. Biol. Chem. **270**(47): 28364-28373.

Feng, X.-H. and Derynck, R. (1997). "A kinase subdomain of transforming growth factor-β (TGF-β) type I receptor determines the TGF-β intracellular signaling specificity." EMBO J **16**(13): 3912-3923.

Ferguson, E. L. and Anderson, K. V. (1992a). "*decapentaplegic* acts as a morphogen to organize dorsal-ventral pattern in the *Drosophila* embryo." Cell **71**: 451-461.

Ferguson, E. L. and Anderson, K. V. (1992b). "Localized enhancement and repression of the activity of the TGF-β family member, *decapentaplegic*, is necessary for dorsal-ventral pattern formation in the *Drosophila* embryo." Development **114**(583-597).

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. and Mello, C. C. (1998). "Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*." Nature **391**: 806-811.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L. and Postlethwait, J. (1999). "Preservation of duplicate genes by complementary, degenerative mutations." Genetics **151**(4): 1531-1545.

François, V., Solloway, M., O'Neill, M., Emery, J. and Bier, E. (1994). "Dorsal-Ventral patterning of the *Drosophila* embryo depends on a putative negative growth factor encoded by the *short gastrulation* gene." Genes Dev. **8**: 2602-2616.

Franzen, P., ten Dijke, P., Ichijo, H., Yamashita, H., Schulz, P., Heldin, C. H. and Miyazono, K. (1993). "Cloning of a TGFβ type I receptor that forms a heteromeric complex with the TGFb type II receptor." Cell **75**: 681-692.

Frasch, M., Hoey, T., Rushlow, C., Doyle, H. and Levine, M. (1987). "Characterzation and localization of the *even-skipped* protein of *Drosophila*." EMBO J **6**(3): 749-759.

Frasch, M. (1995). "Induction of visceral and cardiac mesoderm by ectodermal DPP in the early *Drosophila* embryo." Nature **374**: 464-467.

Galliot, B. (2000). "Conserved and divergent genes in apex and axis development of cnidarians." Curr. Opinion Gen. Dev. **10**: 629-637.

Gardiner, D. M., Blumberg, B., Komine, Y. and Bryant, S. V. (1995). "Regulation of HoxA expression in developing and regenerating axolotl limbs." Development **121**: 1731-1741.

Gauchat, D., Kreger, S., Holstein, T. and Galliot, B. (1998). "*prdl-a*, a gene marker for *Hydra* apical differentiation related to triploblastic paired-like head-specific genes." Development **125**(9): 1637-1645.

Gauchat, D., Mazet, F., Berney, C., Schummer, M., Kreger, S., Pawlowski, J. and Galliot, B. (2000). "Evolution of Antp-class genes and differential expression of *Hydra Hox/para Hox* genes in anterior patterning." PNAS **97**(9): 4493-4498.

Gehring, W. J. (1987). "Homeoboxes in the study of development." Science **236**: 1245-1252.

Gehring, W. J., Affolter, M. and Burglin, T. (1994). "Homeodomain proteins." Annu. Rev. Biochem. **63**: 487-526.

Gelbart, W. M., Irish, V. F., St Johnston, R. D., Hoffmann, F. M., Blackman, R. K., Segal, D., Posakony, L. M. and Grimaila, R. (1985). "The *decapentaplegic* gene complex in *Drosophila melanogaster*." Cold Spring Harbor Symp. quant. Biol. **50**: 119-125.

Gelbart, W. M. (1989). "The *decapentaplegic* gene: a TGF-β homologue controlling pattern formation in *Drosophila*." Development **Supp.**: 65-74.

Georgi, L. L., Albert, P. S. and Riddle, D. L. (1990). "*daf-1*, a *C.elegans* gene controlling dauer larva development, encodes a novel protein receptor kinase." Cell **61**: 635-645.

Gilbert, T. L., Haldeman. B. A., Mulvihill, E, O'Hara, P. J. (1992). "A mammalian homologue of a transcript from the *Drosophila pecanex* locus." J. Neurogenet. **8**(3): 181-187.

Graff, J. M., Bansal, A. and Melton, D. A. (1996). "*Xenopus* Mad proteins transduce distinct subsets of signals for the TGFβ superfamily." Cell **85**: 479-487.

Grasso, L. C., Hayward, D.C., Trueman, J.W.H., Hardie, K.M., Janssens, P.A. and Ball, E.E. (2001). "The evolution of nuclear receptors: evidence from the coral, *Acropora*." Mol. Phylogenet. Evolution **21**: 93-102.

Green, C. R. and Flower, N. E. (1980). "Two new sepate junctions in the phylum Coelenterata." J. Cell. Sci. **42**: 43-59.

Groger, H. and Schmid, V. (2001). "Larval development in Cnidaria: a connection to Bilateria?" **29**(3): 110-4.

Grotzinger, J. P. and Knoll, A.H. (1995). "Anomalous carbonate precipitates: is the Precambrian the key to the Permian?" Palaios. **10**(6): 578-596.

Haerry, T. E., Khalsa, O., O'Connor, M. B. and Wharton, K. A. (1998). "Synergistic signaling by two BMP ligands through the SAX and TKV receptors controls wing growth and patterning in *Drosophila*." Development **125**: 3977-3987.

Hahn, S. A., Schutte, N., Hoque, A. T., Moskaluk, C. A., da Costa, L. T., Rozenblum, E., Weinstein, C. L., Fischer, A., Yeo, C. J., Hruban, R. F. and Kern, S. E. (1996). "DPC4, a candidate tunor suppressor gene at human chromosome 18q21.1." Science **271**: 350-353.

Hanks, S. K. and Quinn, A. M. (1991). "Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members." Methods in Enzymology **200**: 38-62.

Harlow, E. ane Lane, D. (1988). Antibodies: a laboratory manual, New York, Cold Spring Harbour Laboratory.

Harriott, V. J. (1983). "Coral Reefs." **2**: 9-18.

Harvey, R. P. (1996). " NK-2 homeobox genes and heart development." Dev. Biol. **178**(2): 203-216.

Hata, A., Lo, R. S., Wotton, D., Lagna, G. and Massagué, J. (1997). "Mutations increasing autoinhibition inactivate tumor suppressors Smad2 and Smad4." Nature **388**: 82-87.

Hata, A., Lagna, G., Massagué, J. and Hemmati-Brivanlou, A. (1998). "Smad6 inhibits BMP/Smad1 signaling by specifically competing with the Smad4 tumor suppressor." Genes. Dev. **12**: 186-197.

Hawley, S. H. B., Wunnenberg-Stapleton, K., Hashimoto, C., Laurent, M. N., Watabe, T., Blumberg, B. W. and Cho, K. W. Y. (1995). "Disruption of BMP signals in embryonic *Xenopus* ectoderm leads to direct neural induction." Genes. Dev. **9**: 2923-2935.

Hayward, D. C., Catmull, J., Reece-Hoyes, J.S., Berghammer, H., Dodd, H., Hann, S.J., Miller, D.J. and Ball, E.E. (2001). "Gene structure and larval expression of *cnox-2Am* from the coral *Acropora millepora*." Dev. Genes Evol. **211**(1): 10-19.

Hayward, D., Samuel, G., Pontynen, P. C., Catmull, J., Saint, R., Ball, E. E. and Miller, D. J. (submitted)."DPP signalling predates bilaterality: a DPP ortholog from a non-bilateral animal has dorsalising activity in *Drosophila*."

Heldin, C.-H., Miyazono, K., and ten Dijke, P. (1997). "TGF-β signalling from cell membrane to nucleus through SMAD proteins." Nature **390**: 465-471.

Ho, C.-Y., Houart, C., Wilson, S. W. and Stainier, D. Y. R. (1999). "A role for the extraembryonic yolk syncytial layer in patterning the zebrafish embryo suggested by properties of the *hex* gene." Current Biology **9**: 1131-1134.

Hobmayer, E., Hatta, M., Fischer, R., Fujisawa, T., Holstein, T. W. and Sugiyama, T. (1996). "Identification of a *Hydra* homologue of the *beta-catenin/plakoglobin/armadillo* gene family." Gene **172**: 155-159.

Hobmayer, B., Rentzsch, F., Kuhn, K., Happel, C. M., von Laue, C. C., Snyder, P., Rothbacher, U. and Holstein, T. W. (2000). "WNT signalling molecules act in axis formation in the diploblastic metazoan *Hydra*." Nature **407**: 186-188.

Hogan, B. L. M. (1996). "Bone morphogenetic proteins: multifunctional regulators of vertebrate development." Genes. Dev. **10**: 1580-1594.

Holland, P. W. H. (1998). "Major transitions in animal evolution: A developmental genetic perspective." American Zoologist **38**: 829-842.

Holland, P. W. H. (1999). "The future of evolutionary developmental biology." Nature (Supp) **402**: C41-C44.

Holley, S. A., Neul, J. L., Attisano, L., Wrana, J. L., Sasai, Y., O'Connor, M. B., De Robertis, E. M. and Ferguson, E. L. (1996). "The *Xenopus* dorsalizing factor noggin ventralizes *Drosophila* embryos by preventing DPP from activating its receptor." Cell **86**: 607-617.

Hoodless, P. A., Haerry, T, Abdollah, S., Stapleton, M., O'Conner, M. B., Attisano, L. and Wrana, J. L. (1996). "MADR1, a MAD-related protein that functions in BMP2 signaling pathways." Cell **85**: 489-500.

Hoppler, S. and Bienz, M. (1994). "Specification of a single cell type by a *Drosophila* homeotic gene." Cell **76**(4): 689-702.

Horsfield, J., Penton, A., Secombe, J., Hoffman, F.M. and Richardson, H.(1998). "*decapentaplegic* is required for arrest in G1 phase during *Drosophila* eye development." Development **125**(24): 5069-78.

Huitorel, P. (1988). "From cilia and flagella to intracellular motility and back again: a review of a few aspects of microtubule-based motility." Biol. Cell. **63**(2): 249-258.

Imamura, T., Takase, M., Nishihara, A., Oeda, E., Hanai, J-I., Kawabata, M. and Miyazono, K. (1997). "Smad6 inhibits signalling by the TGF-b superfamily." Nature **389**: 622-626.

Immergluck, K. P., Lawrence, P. A. and Bienz, M. (1990). "Induction across germ layers in *Drosophila* mediated by a genetic cascade." Cell **62**: 261-268.

Irish, V. F. and Gelbart, W. M. (1987). "The *decapentaplegic* gene is required for dorsal-ventral patterning of the *Drosophila* embryo." Genes Dev. **1**: 868-879.

Jones, C. M., Broadbent, J., Thomas, P. Q., Smith, J. C. and Beddington, R. S. P. (1999). "An anterior signalling centre in *Xenopus* revealed by the homeobox gene *XHex*." Current Biology **9**: 946-954.

Kato, A., Ozawa, F., Saitoh, Y., Fukazawa, Y., Sugiyama, H. and Inokuchi, K. (1998). "Novel members of the Vesl/Homer family of PDZ proteins that bind metabotropic glutamate receptors." J Biol. Chem. **273**(37): 23969-23975.

Kawabata, M., Inoue, H., Hanyu, A., Imamura, T. and Miyazono, K. (1998). "Smad proteins exist as monomers *in vivo* and undergo homo- and hetero-oligomerization upon activation by serine/threonine kinase receptors." EMBO J **17**(14): 4056-4065.

Keng, V. W., Fujimori, K. E., Myint, Z., Tamamaki, N., Nojyo, Y. and Noguchi, T. (1998). "Expression of *Hex* mRNA in early murine postimplantation embryo development." FEBS Lett. **426**: 183-186.

Keng, V. W., Yagi, H., Ikawa, M., Nagano, T., Myint, Z., Yamada, K., Tanaka, T., Sato, A., Muramatsu, I., Okabe, M., Sato, M., Noguchi, T. (2000). "Homeobox gene Hex is essential for onset of mouse embryonic liver development and differentiation of the monocyte lineage." Biochem. Biophys. Res. Commun. **276**(3): 1155-1161.

Kim, J., Johnson, K., Chen, H-J., Caroll, S. and Laughon, A. (1997). "*Drosophila* Mad binds to DNA and directly mediates activation of *vestigal* by Decapentaplegic." Nature **388**: 304-307.

Kingsley, D. M. (1994). "The TGF-β superfamily: new members, new receptors, and new genetic tests of function in different organisms." Genes. Dev. 8: 133-146.

Kirschner, M. and Gerhart., J. (1998). "Evolvibility." PNAS USA **95**(15): 8420-8427.

Kostoruch, Z., Kostrouchova, M., Love, W., Jannini, E., Piatigorsky, J. and Rall, J. E. (1998). "Retonoic acid X receptor in the diploblast, *Tripedalia cystophora*." PNAS USA **95**(13442-13447).

Kretzschmar, M., Liu, F., Hata, A., Doody, J., and Massagué, J. (1997). "The TGF-β family mediator Smad1 is phosphorylated directly and activated functionally by the BMP receptor kinase." Genes. Dev. **11**: 984-995.

Krishna, S., Maduzia, L. L. and Padgett, R. W. (1999). "Specificity of TGFbeta signaling is conferred by distinct type I receptors and their associated SMAD proteins in *Caenorhabditis elegans*." Development **126**(2): 251-260.

Kurihara, T., Kitamura, K., Takaoka, K. and Nakazato, H. (1993). "Murine Bone Morphogenetic Protein-4 gene: existance of multiple promoters and exons for the 5'-untranslated region." Biochem. Biophys. Res. Commun. **192**(3): 1049-1056.

Kusanagi, K., Inoue, H., Ishidou, Y., Mishima, H. K., Kawabata, M. and Miyazono, K. (2000). "Characterization of a Bone Morphgenetic Protein-responsive Smad-binding element." Mol. Biol. Cell **11**: 555-565.

Labbe, E., Silvestri, C., Hoodless, P. A., Wrana, J. L. and Attisano, L. (1998). "Smad2 and Smad3 positively and negatively regulate TGF β-dependent transcription through the *forkhead* DNA-inding protein FAST2." Mol. Cell. **2**: 109-120.

LaBonne, S.G. and Furst, A. (1989a). "Differentiation in vitro of neural precursor cells from normal and Pecanex mutant *Drosophila* embryos." J. Neurogenet. **5**(2): 99-104.

LaBonne, S. G., Sunitha, I., Mahowald, A. P. (1989b). "Molecular genetics of *pecanex*, a maternal-effect neurogenic locus of *Drosophila melanogaster* that potentially encodes a large transmembrane protein." Dev. Biol. **136**(1): 1-16.

Lagna, G., Hata, A., Hemmati-Brivanlou, A. and Massagué, J. (1996). "Partnership between DPC4 and SMAD proteins in TGF-b signaling pathways." Nature **383**: 832-836.

Lechleider, R. J., de Caestecker, M. P., Dehejia, A., Polymeropoulos, M. H. and Roberts, A. B. (1996). "Serine phosphorylation, chromosomal localization, and Transforming Growth Factor-β signal transduction by human *bsp-1*." J. Biol. Chem. **271**(30): 17617-17620.

Lecuit, T., Brook, W. J., Ng, M., Calleja, M., Sun, H. and Cohen, S. M. (1996). "Two distinct mechanisms for long-range patterning by Decapentaplegic in the *Drosophila* wing." Nature **381**: 387-393.

Lecuit, T. and Cohen, S. M. (1998). "Dpp receptor levels contribute to shaping the Dpp morphogen gradient in the *Drosophila* wing imaginal disc." Development **125**: 4901-4907.

Lee, Y. H., Huang, G.M., Cameron, R.A., Graham, G., Davidson, E.H., Hood, L. and Britten, R.J. (1999). "EST analysis of gene expression in early cleavage-stage sea urchin embryos." Development **126**(17): 3857-3867.

Leimeister C, and Externbrink, A., Klamt B, Gessler M. (1999). "Hey genes: a novel subfamily of hairy- and enhancer of split related genes specifically expressed during mouse embryogenesis." Mech. Dev. **85**(1-2): 173-177.

Lim, R., Miller, J.F. and Zaheer, A. (1989). "Purification and characterization of glia maturation factor beta: a growth regulator for neurons and glia." PNAS USA. **86**(10): 3901-3905.

Lin, H. Y., Wang, X. F., Ng-Eaton, E., Weinberg, R. A. and Lodish, H. F. (1992). "Expression Cloning of the TGF-β type II receptor, a functional transmembrane serine/treonine kinase." Cell **68**: 1-20.

Liu, F., Ventura, F., Doody, J. and Massagué, J. (1995). "Human Type II receptor for Bone Morphogenic Proteins (BMP): Extension of the two-kinase Receptor model to the BMPs." Mol. Cell. Biol. **15**(7): 3479-3486.

Liu, F., Hata, A., Baker, J. C., Doody, J., Carcamo, J., Harland, R. M. and Massague, J. (1996). "A human Mad protein acting as a BMP-regulated transcriptional activator." Nature **381**: 622-623.

Liu, F., Pouponnot, C. and Massagué, J. (1997). "Dual role of the Smad4/DPC4 tumor suppressor in TGFbeta-inducible transcriptional complexes." Genes. Dev. **11**(23): 3157-3167.

Lo, R. S., Chen, Y-G., Shi, Y., Pavletich, N. P. and Massagué, J. (1998). "The L3 loop: a structural motif determining specific interactions between SMAD proteins and TGF-β receptors." EMBO J **17**(4): 996-1005.

Macias-Silva, M., Abdollah, S., Hoodless, P. A., Pirone, R., Attisano, L. and Wrana, J. L. (1996). "MADR2 Is a substate of the TGFβ receptor and its phosphorylation is required for nuclear accumulation and signaling." Cell **87**: 1215-1224.

Maggert, K., Levine, M and Frasch, M. (1995). "The somatic-visceral subdivision of the embryonic mesoderm is initiated by dorsal gradient thresholds in *Drosophila*." Development **121**(7): 2107-2116.

Manak, J. R., Mathies, L. and Scott, M. (1995). "Regulation of a *decapentaplegic* midgut enhancer by homeotic proteins." Development **120**: 3605-3619.

Marqués, G., Musacchio, M., Shimell, M. J., Wunnenberg-Stapeton, K., Cho, K. W. Y. and O'Connor, M. B. (1997). "Production of a DPP Activity Gradient in the Early *Drosophila* Embryo through the Opposing Actions of the SOG and TLD Proteins." Cell **91**: 417-426.

Martinez Barbera, J. P., Clements, M., Thomas, P., Rodriguez, T., Meloy, D., Kioussis, D. and Beddington, R. S. P. (2000). "The homeobox gene *Hex* is required in definitive endodermal tissues for normal forebrain, liver and thyroid formation." Development **127**: 2433-2445.

Martinez Barbera, J. P. and Beddington, R. S. P. (2001). "Getting your head around Hex and Hesx1: Forebrain formation in mouse." Dev. Biol. **45**: 327-336.

Marty, T., Vigano, M.A., Ribeiro, C., Nussbaumer, U., Grieder, N.C. and Affolter, M. (2001). " A HOX complex, a repressor element and a 50 bp sequence confer regional specificity to a DPP-responsive enhancer." Development **128**(14): 2833-2845.

Massagué, J. (1990). "The transforming growth factor-β family." Annu. Rev. Cell Biol. **6**: 597-641.

Massagué, J. (1992). "Receptors for the TGF-β Family." Cell **69**: 1067-1070.

Massagué, J. (1998). "TGF-β signal transduction." Annu. Rev. Biocham. **67**: 753-791.

Massagué, J. and Wotton, D. (2000). "Transcriptional control by the TGF-beta/Smad signaling system." EMBO J **19**(8): 1745-1754.

Mathews, L. S. and Vale, W. W. (1991). "Expression cloning of an activin receptor, a predicted transmembrane serine kinase." Cell **65**: 973-982.

Meersseman, G., Verschueren, K., Nelles, L., Blumenstock, C., Kraft, H., Wuytens, G., Remacle, J., Kozak, C. A., Tylzanowski, P., Niehrs, C. and Huylebroeck, D. (117). "The C-terminal domain of Mad-like signal transducers is sufficient for biological activity in the *Xenopus* embryo and transcriptional activation." Mech. Dev. **61**(1-2): 127-140.

Miller, D. J. and Ball, E. E. (2000). "The coral *Acropora*: what it can contribute to our knowledge of metazoan evolution and the evolution of developmental processes." Bioessays **22**(3): 291-296.

Miller, D. J. and Harrison, P. (1990). "Molecular and developmental biology of the Cnidaria-basic aspects and phylogenetic implications." Aust. J. Biotechnol. **4**(4): 241-245.

Miller, D. J. and Miles, A. (1993). "Homeobox genes and the zootype." Nature **365**: 215-216.

Mochizuki, K., Sano, H., Kobayashi, S., Nishimiya-Fujisawa, C. and Fujisawa, T. (2000). "Expression and evolutionary conservation of *nanos*-related genes in *Hydra*." Dev. Genes Evol. **210**(12): 591-602.

Mornet D, and Bonet-Kerrache, A. (2001). "Neurocalcin-actin interaction." Biochim. Biophys. Acta. **1549**(2): 197-203.

Muramatsu, M., Yan, J., Eto, K., Tomoda, T., Yamada, R. and Arai, K. (1997). "A chimeric serine/threonine kinase receptor system reveals the potential of multiple type 11 receptors to cooperate with transforming growth factor-β type I receptor." Mol. Biol. Cell **8**: 469-480.

Nakamura, T. Y., Pountney, D. J., Ozaita, A., Nandi, S., Ueda, S., Rudy, B. and Coetzee, W.A. (2001). "A role for frequenin, a Ca^{2+}-binding protein, as a regulator of Kv4 K^+-currents." PNAS USA. **98**(22): 12808-12813.

Nakao, A., Imamura, S. S., Kawabata, M., Ishisaki, A., Oeda, E., Tamaki, K., Hanai, J., Heldin, C-H., Miyazono, K. and ten Dijke, P. (1997a). "TGF-b receptor-mediated signalling through Smad2, Smad3 and Smad4." EMBO Journal **16**(17): 5353-5362.

Nakao, A., Roijer, E., Imamura, T., Souchelnytskyi, S., Stenman, G., Heldin, C-H. and ten Dijke, P. (1997b). "Identification of Smad2, a human-related protein in the transforming growth factor β signaling pathway." J. Biol. Chem. **272**(5): 2896-2900.

Nakao, A., Afrakhte, M., Moren, A., Nakayama, T., Christian, J. L., Heuchel, R., Itoh, S., Kawabata, M., Heldin, N-E., Heldin, C-H. and ten Dijke, P. (1997c). "Identification of Smad7, a TGFβ-inducible antagonist of TGFb signalling." Nature **389**(631-635).

Nellen, D., Burke, R., Struhl, G. and Basler, K (1996). "Direct and long-range action of a DPP morphogen gradient." Cell **85**: 357-368.

Newfeld, S. J., Padgett, R.W., Findley, S.D., Richter, B.G., Sanicola, M., de Cuevas, M. and Gelbart, W.M. (1997). "Molecular evolution at the *decapentaplegic* locus in *Drosophila*." Genetics **145**(2): 297-309.

Newfeld, S. J., Wisotzkey, R. G. and Kumar, S. (1999). "Molecular evolution of a developmental pathway: phylogenetic analyses of transforming growth factor-β family ligands, receptors and smad signal transducers." Genetics **152**: 783-795.

Nielsen, C. (1995). Animal Evolution: interrelationships of the living phyla. Oxford Univ. press.

Nishitoh, H., Ichijo, H., Kimura, M., Matsumoto, T., Makishima, F., Yamaguchi, A., Yamashita, H., Enomoto, S. and Miyazono, K. (1996). "Identification of Type I and Type II serine/threonine kinase receptors for Growth/Differentiation Factor-5." J. Biol. Chem. **271**: 21345-21352.

Nohno, T., Ishikawa, T., Saito, T., Hosokawa, K., Noji, S., Wolsing, D-H. and Rosenbaum, J. S. (1995). "Identification of a human Type II Receptor for Bone Morphogenetic Protein-4 that forms differential heteromeric complexes with Bone Morphogenetic Protein type I receptors." J. Biol. Chem. **270**(38): 22522-22526.

Padgett, R. W., St Johnston, R. D. and Gelbart, W. M. (1987). "A transcript from a *Drosophila* pattern gene predicts a protein homologous to the transforming growth factor-beta family." Nature **325**(6099): 81-84.

Padgett, R. W., Wozney, J. M., Gelbart, W. M. (1993). "Human BMP sequences can confer normal dorsal-ventral patterning in the *Drosophila* embryo." PNAS USA **90**: 2905-2909.

Panganiban, G. E. F., Reuter, R., Scott, M. P. and Hoffmann, F. M. (1990). "A *Drosophila* growth factor homolog, *decapentaplegic*, regulates homeotic gene expression within and across germ layers during midgut morphogenesis." Development **110**: 1041-1050.

Pellizzari, L., D'Elia, A., Rushtighi, A., Manfioletti, G., Tell, G. and Damante, G. (2000). "Expression and function of the homeodomain-containing protein Hex in thyroid cells." Nucl. Acids. Res. **28**(13): 2503-2511.

Pőpperl, H., Schmidt, C., Wilson, V., Hume, C., Dodd, J., Krumlauf, R. and Beddington, R. S. P. (1997). "Misexpression of *Cwnt8C* in the mouse induces an ectopic embryonic axis and causes a truncation of the anterior neurectoderm." Development **124**: 2997-3005.

Posakony, L. G., Raftery, L. A. and Gelbart, W. M. (1991). "Wing formation in *Drosophila melanogaster* requires *decapentaplegic* function along the anterior-posterior compartment boundary." Mech. Dev. **33**: 69-82.

Raftery, L. A., Twombly, V., Wharton, K. and Gelbart, W. M. (1995). "Genetic screens to identify elements of the *decapentaplegic* signaling pathway in *Drosophila*." Genetics **139**: 241-154.

Ray, R. P., Arora, K., Nusslein-Volhard, C. and Gelbart, W.M. (1991). "The control of cell fate along the dorsal-ventral axis of the *Drosophila* embryo." Development **113**(1): 35-54.

Richter, B., Long, M., Lewontin, R.C. and Nitasaka, E. (1997). "Nucleotide variation and conservation at the *dpp* locus, a gene controlling early development in *Drosophila*." Genetics **145**(2): 311-323.

Rosenzweig, B. L., Imamura, T., Okadome, T., Cox, G. N., Yamashita, H., ten Dijke, P., Heldin, C-H. and Miyazono, K. (1995). "Cloning and characterization of a human type II receptor for bone morphogenetic proteins." PNAS **92**: 7632-7636.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular cloning: a laboratory manual, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, N.Y.

Sampath, T., Rashka, K., Doctor, J., Tucker, R. and Hoffmann, F. M. (1993). "*Drosophila* TGF-β superfamily proteins induce endochondrial bone formation in mammals." PNAS USA **90**: 6004-6008.

Sasai, Y., Lu, B., Steinbeisser, H. and de Robertis, E. M. (1995). "Regulation of neural induction by the *chd* and BMP-4 antagonistic patterning signals in *Xenopus*." Nature **376**: 333-336.

Savage, C., Das, P., Finelli, A. L., Townsend, S. R., Sun, C., Baird, S. E. and Padgett, R. W. (1996). "*Caenorhabditis elegans* genes *sma-2, sma-3*, and *sma-4* define a conserved family of transforming growth factor β pathway components." PNAS, USA **93**: 790-794.

Schierwater, B., Murtha, M., Dick, M., Ruddle, F.H. and Buss, L.W. (1991). "Homeoboxes in cnidarians." J. Exp. Zool. **260**: 413-416.

Schmidt, J. E., Suzuki, A., Ueno, N. and Kimelman, D. (1995). "Localized Bmp-4 mediates dorsal/ventral patterning in the early *Xenopus* embryo." Dev. Biol. **169**: 37-50.

Schummer, M., Scheurlen, I., Schaller, C. and Galliot, B.. (1992). "HOM/HOX homeobox genes are present in *Hydra* (*Chlorohydra viridissima*) and are differentially expressed during regeneration." EMBO J **11**(5): 1815-1823.

Segal, D. and Gelbart, W. M. (1985). "Shortvein, A new component of the *decapentaplegic* gene complex in *Drosophila melanogaster*." Genetics **109**(1): 119-143.

Sekelsky, J. J., Newfeld, S. J., Raftery, L. A., Chartoff, E. H. and Gelbart, W. M. (1995). "Genetic Characterization and Cloning of *Mothers against dpp,* a gene required for *decapentaplegic* function in *Drosophila melanogaster*." Genetics **139**: 1347-1358.

Shawlot, W., Wakamiya, M., Kwan, K-M., Kania, A., Jessell, T. and Behringer, R. (1999). "Lim1 is required in both primitive streak-derived tissues and visceral endoderm for head formation in the mouse." Development **126**: 4925-4932.

Shi, Y., Hata, A., Lo, R. S., Massagué, J. and Pavletich, N. P. (1997). "A structural basis for mutational inactivation of the tumor suppresor Smad4." Nature **388**: 87-93.

Shi, Y., Wang, Y-F., Jayaraman, L., Yang, H., Massagué, J. amd Pavetich, N. P. (1998). "Crystal Structure of a Smad MH1 Domain Bound to DNA: Insights on DNA Binding in TGF-b Signaling." Cell **94**: 585-594.

Shi, X., Yang, X., Chen, D., Chang, Z. and Cao, X. (1999). "Smad1 interacts with homeobox DNA-binding proteins in bone morphogenetic protein signaling." J. Biol. Chem. **274**: 13711-13717.

Slack, J. M. W., Holland, P. W. H. and Graham, C. F. (1993). "The zootype and the phylotypic stage." Nature **361**: 490-492.

Souchelnytskyi, S., Tamaki, K., Engstrom, U., Wernstedt, C., ten Dijke, P., and Heldin, C. H. (1997). "Phosphorylation of Ser465 and Ser467in the C terminus of Smad2 mediates interactionwith Smad4 and is required for transforming growth factor-β signaling." J. Biol. Chem. **272**: 28107-28115.

St. Johnston, D. D. and Gelbart, W. M. (1987). "Decapentaplegic transcripts are localized along the dorsal-ventral axis of the *Drosophila* embryo." EMBO J **6**(9): 2785-2791.

St. Johnston, R. D., Hoffmann, F. M., Blackman, R., Segal, D. and Grimaila, R. (1990). "Molecular organization of the *decapentaplegic* gene in *Drosophila melanogaster*." Genes. Dev. **4**: 1114-1127.

Staehling-Hampton, K., Hoffmann, F. M., Baylies, M. K., Rushton, E. and Bate, M. (1994). "*dpp* induces mesodermal gene expression in *Drosophila*." Nature **372**: 783-786.

Steidl C, Leimeister, C., Klamt B, Maier M, Nanda I, Dixon M, Clarke R, Schmid M, Gessler M. (2000). "Characterization of the human and mouse HEY1, HEY2, and HEYL genes: cloning, mapping, and mutation screening of a new bHLH gene family." Genomics. **66**(2): 195-203.

Suga, H., Ono, K. and Miyata, T. (1999). "Multiple TGF-beta receptor related genes in sponge and ancient gene duplications before the parazoan-eumetazoan split." FEBS Lett. **453**(3): 346-350.

Summerton, J. and Weller, D. (1997). "Morpholino antisense oligomers: design, preparation,and properties." Antisense Nucleic Acid Drug Dev. **7**: 187–195.

Suzuki, A., Thies, R. S., Yamaji, N., Song, J. J., Wozney, J. M., Murakami, K. and Ueno, N. (1994). "A truncated bone morphogenetic protein receptor affects dorsal-ventral patterning in the early *Xenopus* embryo." PNAS USA **91**: 10255-10259.

Suzuki, M. M. and Satoh, N. (2000). "Genes expressed in the amphioxus notochord revealed by EST analysis." Dev. Biol. **224**: 168-177.

Swofford, D. L. (2000). PAUP* (Phylogenetic analysis using parsimony, and other methods), version 4b3.

Szent-Gyorgyi, A. G. (1975). "Calcium regulation of muscle contraction." Biophys. J. **15**(7): 707-23.

Takase, M., Imamura, T., Sampath, T. K., Takeda, K., Ichijo, H., Miyazono, K. and Kawabata, M. (1998). "Induction of Smad6 mRNA by bone morphogenetic proteins." Biochem. Biophys. Res. Commun. **244**: 26-29.

Tam, P. P. and Steiner, K. A. (1999). "Anterior patterning by synergistic activity of the early gastrula organizer and the anterior germ layer tissues of the mouse embryo." Development **126**: 5171-5179.

Tanaka, T., Inazu, T., Yamada, K., Myint, Z., Keng., V. W., Inoue, Y., Taniguchi, N. and Noguchi, T. (1999). "cDNA cloning and expression of rat homeobox gene, *Hex*, and functional characterization of the protein." Biochem. J. **339**: 111-117.

ten Dijke, P., Yamashita, H., Ichijo, H., Franzen, P., Lahio, M., Miyazono, K. and Heldin, C-H. (1994). "Characterization of type I receptors for transforming growth factor-β and activin." Science **264**: 101-104.

Thomas, P. Q. and Beddington, R. S. P. (1996). "Anterior primitive endoderm may be responsible for patterning the anterior neural plate in the mouse embryo." Curr. Biol. **6**: 1487-1496.

Thomas, P. Q., Brickman, J., Popperl, H., Krumlauf, R. and Beddington, R. S. P. (1997). "Axis duplication and anterior identity in the mouse embryo." Symposia of Quantitative Biology, Cold Spring Harbour Press.

Thomas, P. Q., Brown, A. and Beddington, R. S. P. (1998). "*Hex:* a homeobox gene revealing peri-implantation asymmetry in the mouse embryo and an early transient marker of endothelial cell precursors." Development **125**: 85-94.

Thomsen, G. H. (1996). "*Xenopus mothers against decapentaplegic* is an embryonic ventralizing agent that acts downstream of the BMP-2/4 receptor." Development **122**: 2359-2366.

Tsuneizuni, K., Nakayama, T., Kamoshida, Y., Kornberg, T., Christian, J. and Tabata, T. (1997). "*Daughters against dpp* modulates *dpp* organizing activity in *Drosophila* wing development." Nature **389**: 627-630.

Valentine, J. W. (1994). "Late Precambrian bilaterians: grades and clades." PNAS USA. **91**(15): 6751-6757.

Valentine, J. W., Jablonski, D. and Erwin, D. H. (1999). "Fossils, molecules and embryos: new perspectives on the Cambrian explosion." Development **126**: 851-859.

van den Wijngaard, A., van Kraay, M., van Zoelen, E. J. J., OLijve, W. and Boersma, C. J. C. (1996). "Genomic organization of the human Bone Morphogenetic Protein-4 gene: Molecular basis for multiple transcripts." Biocem. Biophys. Res. Commun. **219**: 789-794.

Vandermeulen, J. H. (1974). "Studies on reef corals. II. Fine structure of planktonic planula larva of *Pocillopora damicornia*, with emphasis on the aboral epidermis." Marine Biology **27**: 239-249.

Vandermeulen, J. H. (1975). "Studies on reef corals. III. Fine structural changes of calicoblast cells in *Pocillopora damicornis* during settling and calcification." Marine Biology **31**: 69-77.

Varlet, I., Collignon, J. and Robertson, E. J. (1997). "*nodal* expression in the primitive endoderm is required for specification of the anterior axis during mouse gastrulation." Development **124**: 1033-1044.

Venot, C., Maratrat, M., Dureuil, C., Conseiller, E., Bracco, L. and Debussche, L. (1998). "The requirement for the p53 proline-rich functional domain for mediation of apoptosis is correlated with specific PIG3 gene transactivation and with transcriptional repression." EMBO J **17**: 4668-4679.

Vibede, N., Hauser, F., Williamson, M. and Grimmelikhuijzen, C. J. (1998). "Genomic organization of a receptr from sea anemones, structurally and eolutionarily related to glycoprotein hormone receptors from mammals." Biocem. Biophys. Res. Commun. **252**: 497-501.

Vignal, E., Blangy, A., Martin, M., Gauthier-Rouviere, C. and Fort, P. (2001). "kinectin is a key effector of RhoG microtubule-dependent cellular activity." Mol. Cell. Biol. **21**(23): 8022-8034.

Wan, K. L., Blackwell, J.M. and Ajioka, J.W. (1996). "*Toxoplasma gondii* expressed sequence tags: insight into tachyzoite gene expression." Mol. Biochem. Parasitol. **75**(2): 179-186.

Wharton, K. A., Ray, R. P. and Gelbart, W. M. (1993). "An activity gradient of *decapentaplegic* is necessary for the specification of dorsal pattern elements in the *Drosophila* embryo." Development **117**: 807-822.

Willmer, P. (1990). Invertebrate Relationships. Patterns in Animal Evolution, Cambridge Univ. Press, Cambridge, UK.: 188-191.

Wilson, P. A. and Hemmati-Brivanlou, A. (1995). "Induction of epidermis and inhibition of neural fate by Bmp-4." Nature **376**: 331-333.

Wisotzkey, R. G., Mehra, A., Sutherland, D., Dobens, L. L., Dohrmann, C., Attisano, L., and Raftery, L. A. (1998). "*medea* is a *Drosophila smad4* homolog that is differentially required to potentiate DPP responses." Development **125**: 1433-1445.

Wrana, J. L., Attisano, L., Carcamo, J., Zentella, A., Doody, J., Lahio, M., Wang, X-F. and Massagué, J. (1992). "TGFβ signals through a heteromeric protein kinase receptor complex." Cell **71**: 1003-1014.

Wrana, J. L., Atissano, L., Wieser, R., Ventura, F and Massagué, J. (1994). "Mechanism of activation of the TGF-β Receptor." Nature **370**: 341-347.

Wu, R.-Y., Zhang, Y., Feng, X-H. and Derynck, R. (1997). "Heteromeric and homomeric interactions correlate with signaling activity and functional cooperativity of Smad3 and Smad4/DPC4." Mol. Cell. Biol **17**(5): 2521-2528.

Xu, R. J., Kim, J., Taira, M., Zhan, S., Sredni, D. and Kung, H. (1995). "A dominant-negative bone morphogenetic protein-4 receptor causes neuralization in *Xenopus* ectoderm." Biochem. Biophys. Res. Commun. **212**: 212-219.

Xu, X., Yin, Z., Hudson, J., Ferguson, E. and Frasch, M. (1998). "Smad proteins act in combination with synergistic and antagonistic regulators to target Dpp responses to the *Drosophila* mesoderm." Genes. Dev. **12**: 2354-2370.

Yamashita, H., ten Dijke, P., Franzen, P., Miyazono, K. and Heldin C-H. (1994). "Formation of hetero-oligomeric complexes of type I and type II receptors for transforming growth factor-beta." J. Biol. Chem. **269**: 20172-20178.

Zawel, L., Dai, J. L., Buckhaults, P., Zhou, S., KInzler, K. W., Volgelstein, B. and Kern, S. E. (1998). "Human Smad3 and Smad4 are sequence-specific transcription activators." Mol. Cell. **1**: 611-617.

Zecca, M., Basler, K. and Struhl, G. (1995). "Sequential organizing activities of *engrailed, hedgehog* and *decapentaplegic* in the *Drosophila* wing." Development **121**: 2265-2278.

Zhang, Y., Feng, X-H., Wu, R-Y. and Derynck, R. (1996). "Receptor-associated Mad homologues synergize as effectors of the TGF-b response." Nature **383**: 168-172.

Zhang, Y., Feng, X. H. and Derynck, R. (1998). "Smad3 and Smad4 cooperate with c-Jun/c-Fos to mediate TGFβ-induced transcription." Nature **394**: 909-913.

Zhang, Y. and Derynck, R. (1999). "Regulation of Smad signalling by protein associations and signalling crosstalk." Trends Cell Biol. **9**(7): 274-279.

Zhou, S., Zawel, L., Lengauer, C., KInzler, K. W. and Vogelstein, B. (1998). "Characterization of human FAST-1, a TGFβ and activin signal transducer." Mol. Cell **2**: 121-127.

Zimmerman, L. B., Jose, M., De Jesus-Escobar, and Harland, R. M. (1996). "The spemann organizer signal Noggin binds and inactivates Bone Morphogenetic Protein 4." Cell **86**: 599-606.

Zusman, S. B., Sweeton, D. and Wieschaus, E. F. (1988). "*short gastrulation*, a mutation causing delays in stage specific cell shape changes during gastrulation in *Drosophila melanogaster*." Dev. Biol. **129**: 417-427.

ADDENDUM

Mis-spelt citations:

Page 7 : Kostoruch *et al.*, 1998

Page 17: Arora and Nusslein-Volhard, 1992; Ferguson
Campos-Ortega and Vassin, 1985.

Page 18: Bryant, 1988

Page 21: Van den Wijngaard et al., 1996

Page 67: Altschul *et al.*, 1997

Altered citations:

Page 1: Campbell, 1993
Corrected to Valentine *et al.*, 1999
Page 3 Campbell, 1993
Corrected to Lundin, 1999
Page 4 Campbell, 1993
Corrected to Lundin, 1999

Page 20, 3rd paragraph, last sentence corrected to:

However, although tempting, it is important not to draw detailed analogies between vertebrate and arthropod dorsal/ventral organisation, as vertebrates have many more additional factors involved in their patterning.
Reference (Holland, 1998)

Page 20, 4th paragraph corrected to:

BMP2, a paralog of BMP4, is also an ortholog of DPP.

Section 6.3.2, 2nd sentence corrected to:

Further, although the entire coding region of *Hydra hex* is still unknown, it is likely to have a sequence length equivalent to that of Hex-*Am* as the *Hydra* Hex C-terminal, similar to Hex-*Am*, is significantly shorter than C-terminal sequences found in higher metazoans (Gauchat *et al.* 2000)

Page 84, 3rd paragraph corrected to:

The early endoderm-specific expression of *hex* in higher metazoans hinted that *hex-Am* may be activated at an early stage of *A. millepora* development, possibly around the time of gastrulation and subsequent formation of the endoderm (24-hour-old embryos; see Section 1.4.1).

Additional reference added:

Lundin, L-G. (1999). "Gene duplications in early metazoan evolution." Cell and Developmental biology 10: 523-530