2003

Isolation and characterisation of a pax group III gene from the freshwater crayfish cherax destructor: Implications for the evolution of muscle regeneration

Robert B. White
Edith Cowan University

Edith Cowan University

Copyright Warning

ISOLATION AND CHARACTERISATION OF A *PAX* GROUP III GENE FROM THE FRESHWATER CRAYFISH *CHERAX DESTRUCTOR*: IMPLICATIONS FOR THE EVOLUTION OF MUSCLE REGENERATION

Robert Bruce White

0961371

This thesis is submitted as partial fulfilment for the award of

Bachelor of Science (Biological Science) Honours

School of Natural Sciences

Edith Cowan University

19 December 2003

USE OF THESIS

The Use of Thesis statement is not included in this version of the thesis.

ABSTRACT

Pax genes encode transcription factors that are highly evolutionarily conserved and are vital for animal development. Vertebrate *Pax* group III genes *Pax3* and *Pax7* are required for proper development of muscular and central nervous systems. In their roles in muscular systems, *Pax3* and *Pax7* specify myoblasts (muscle progenitor cells). *Pax3* is predominantly involved in embryonic myogenesis and specifies embryonic myoblasts. In contrast, *Pax7* specifies adult myoblasts (muscle satellite cells), and has been demonstrated to be vital for adult muscle regeneration. Recent evidence also implicates a role for *Pax7* in the conversion of certain stem cells to the myogenic lineage.

The Australian freshwater crayfish *Cherax destructor* (yabby) possesses a phenomenal capability for almost limitless adult muscle regeneration. Interestingly, *C. destructor* undergoes two distinct types of adult muscle regeneration, and these appear to be driven by two distinct types of muscle progenitor/stem cell. *C. destructor* has a highly regulated system for normal muscle growth, which is strictly regulated around the periodic shedding of the growth-restricting exoskeleton. The muscle regeneration that occurs in this normal growth is driven predominantly by endogenous muscle satellite cells. *C. destructor* also possesses the fascinating ability to regenerate entire limbs, including all of the muscle contained within them. This process appears to utilise circulating haematopoietic stem cells as the dominant muscle progenitor cell (myoblast).

The role for *Pax7* in adult muscle regeneration in vertebrates indicates that *Pax* group III genes may be appropriate candidates to study in muscle regeneration in *C. destructor*. *Pax* group III genes have been isolated in fruit fly, grasshopper, mite and jellyfish, and have not been demonstrated to be expressed during embryonic or adult myogenesis in any of these animals. Prior to this research, *Pax* genes had not been isolated from Crustacea.

I have designed nested partially-degenerate primers to complement conserved regions of known arthropod *Pax* group III gene sequences, and used PCR and RT-PCR of *C. destructor* genomic DNA and embryonic RNA, respectively. Sequencing of reaction products confirmed the presence of a *Pax* group III gene in Crustacea. In this research I have isolated a single *Pax* group III gene from *C. destructor* which I have designated

Cdpax3/7. *Cdpax3/7* is unequivocally a *Pax* group III gene, and appears to be the only *Pax* group III gene present in *C. destructor*. Cdpax3/7 contains two DNA binding domains characteristic of Pax group III proteins, the paired domain and the homeodomain. Cdpax3/7 also contains an octapeptide motif characteristic of the majority of Pax group III proteins. Interestingly, *Cdpax3/7* is expressed as two alternate transcripts. One alternate transcript lacks a 93 nucleotide section corresponding to approximately one third of the paired domain, indicating that it may have altered DNA binding, and therefore function.

RT-PCR expression assays indicate that *Cdpax3/7* is expressed embryonically (in whole *C. destructor* embryos), in normal non-regenerating adult muscle, in adult muscle undergoing normal regeneration, as well as in limb regeneration (both before and after myogenic differentiation). The gross expression pattern of *Cdpax3/7* in adult muscle regeneration is characteristic of the expression patterns of vertebrate *Pax7* in the analogous vertebrate process.

COPYRIGHT AND ACCESS DECLARATION

I certify that this thesis does not, to the best of my knowledge and belief:

(i) incorporate without acknowledgement any material previously submitted for a degree or diploma in any institution of higher education;

(ii) contain any material previously published or written by another person except where due reference is made in the text; or

(iii) contain any defamatory material

Robert White
0961371

19 December 2003

ACKNOWLEDGEMENTS

I wish to extend my sincere gratitude to my inspirational supervisor Dr Annette Koenders, without whom this thesis would not exist. For her limitless and unwavering support and encouragement, her selfless unending contribution, and her knowledge and creativity, and for just knowing exactly when to do all of the right things, I am eternally grateful.

To my wonderful co-supervisor, Dr Mel Ziman, thank you for your guidance and support throughout the year. Thanks for always being there with your door open whenever I needed to vent a disappointment, share an achievement, or just bounce an idea.

I wish to express a huge debt of gratitude to Tina Lamey for massive amounts of assistance in laboratory technique. Thanks Tina, your input was above and beyond the call of duty and I really appreciate it. Thanks to Assoc. Prof. A. Kinnear and Derek Swarts for assistance with microscopy. Thanks to all of the biology and biomedical sciences laboratory staff for assistance with a multitude of harebrained schemes (some of which actually worked!) for setting up yabbies and fruit flies for breeding and embryo harvesting.

A huge thankyou to Gemma Cole for taking on the unenviable task of maintaining my sanity over the year (not quite sure whether she succeeded). Thanks to Julia Wilson, Meghan Thomas, and all of my colleagues doing honours this year for making Uni a very enjoyable place to be. Thanks to my friends for their acceptance and to my parents for their endless support.

Finally, I would like to thank yabbies for involuntarily giving up their limbs in the name of science.

"Nothing in biology makes sense....., except in the light of evolution"

Theodosius Dobzhansky

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BLAST	basic local alignment search tool
BMP	bone morphogenic protein
bp	base pairs
cDNA	complementary deoxyribonucleic acid
CNS	central nervous system
Cdpax3/7	*Cherax destructor* Pax protein similar to Pax3 and Pax7
DEPC	diethyl pyrocarbonate
diH_2O	deionised water
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleotide triphosphate
ECU	Edith Cowan University
EDTA	ethylenediaminetetra-acetate
GC	guanine and cytosine
Gsb	Gooseberry (*D. melanogaster*)
Gsb-n	Gooseberry-neuro (*D. melanogaster*)
HCl	hydrochloric acid
HD	homeodomain
HSC	haematopoietic stem cell
kb	kilobases
Mef2	myocyte enhancer factor 2
$MgCl_2$	magnesium chloride
OCL	ocular carapace length
Pax	paired box
Pby1	Pairberry 1 (*S. americana*)
Pby2	Pairberry 2 (*S. americana*)
PCR	polymerase chain reaction
PD	paired domain
pgIII	Pax group III
PHYLIP	phylogeny inference package
Prd	Paired (*D. melanogaster*)

RNA	ribonucleic acid
RNase	ribonuclease
RT-PCR	reverse transcriptase-polymerase chain reaction
Shh	Sonic hedgehog
SP	side population
Tupax3/7	*Tetranychus urticae* Pax protein similar to Pax3 and Pax7
UV	ultra violet
Wnt	Wingless-int

N.B. the following conventional notation is used;

Pax indicates gene,

Pax indicates protein,

PAX indicates genes from *Homo sapiens*,

PAX indicates protein from *H. sapiens*.

IUPAC Degenerate Nucleotide Code;

Y = C or T	R = A or G
W = A or T	S = C or G
K = T or G	M = C or A
B = C, G or T	D = A, G or T
H = A, C or T	V = A, C or G
N = A, C, G or T	

A = Adenine; T = Thymine; G = Guanine; C = Cytosine

1.0 Introduction

Decapod crustaceans such as the Australian yabby *Cherax destructor* have a phenomenal capability for almost limitless muscle regeneration as an adult, giving them a formidable advantage over potential competitors. The adult myogenesis present in *C. destructor* is far advanced over most other species, in that they can regenerate entire limbs including all the muscle contained within them. The genetic hierarchies responsible for this process are unknown; however, it has been found in many other species that the genetic mechanisms responsible for adult regeneration are often a recapitulation of mechanisms involved in embryonic development (Carlson, 2003; Polesskaya *et al.*, 2003; Muller *et al.*, 2003). Many of the genetic mechanisms that are responsible for embryonic muscle development are highly conserved, with similar genes being involved in similar processes in vertebrates, fruit fly, and even jellyfish (Muller *et al.*, 2003; Baylies & Michelson, 2001). Pax group III transcription factors have been found to specify myoblasts in higher vertebrates, with Pax3 specifying embryonic myoblasts (Goulding *et al.*, 1994) and Pax7 specifying adult myoblasts (muscle satellite cells) (Seale *et al.*, 2000). The sum of recent evidence also implicates Pax7 as playing a role in the conversion of stem cells to the myogenic lineage, introducing an exciting area of research into the use of this transcription factor for clinical treatment of degenerative muscular conditions such as Duchenne Muscular Dystrophy (Parker *et al.*, 2003). *Pax* group III genes have been isolated in chordates, arthropods and jellyfish, and thus far have been found to have a role in myogenesis only in chordates. The purpose of this research was to isolate *Pax* group III genes from *C. destructor* and to determine whether these transcription factors may have a role in adult muscle regeneration in Crustacea.

1.1 The *Pax* Family of Genes

Members of the *Pax* family of genes encode transcription factors that are important regulators of tissue specific developmental processes in animals. Pax proteins contain a highly conserved 128 amino acid DNA binding motif called the paired domain (PD; encoded by a DNA region called the paired box, hence *Pax*). The PD is comprised of a 'pair' of helix-turn-helix motifs, each consisting of three α-helices. In this bipartite structure, the N-terminal motif has been labelled the PAI domain, and the C-terminal motif

has been labelled the RED domain (PAI + RED = PD; Jun & Desplan, 1996). In addition to this, certain groups of Pax proteins possess an equally conserved 60 amino acid DNA binding homeodomain (HD). The HD is structurally similar to the PD domain motifs, in that it consists of three α-helices in a helix-turn-helix motif. Pax proteins also possess a C-terminal transactivation domain, through which transcription of the bound target gene is facilitated (Schafer *et al.*, 1994). Many Pax proteins also contain a conserved octapeptide motif that is believed to be involved in the negative modulation of transcriptional activation (Lechner & Dressler, 1996). The spatial orientation of these domains is depicted in the schematic representation of a Pax group III protein binding DNA in Figure 1.1.

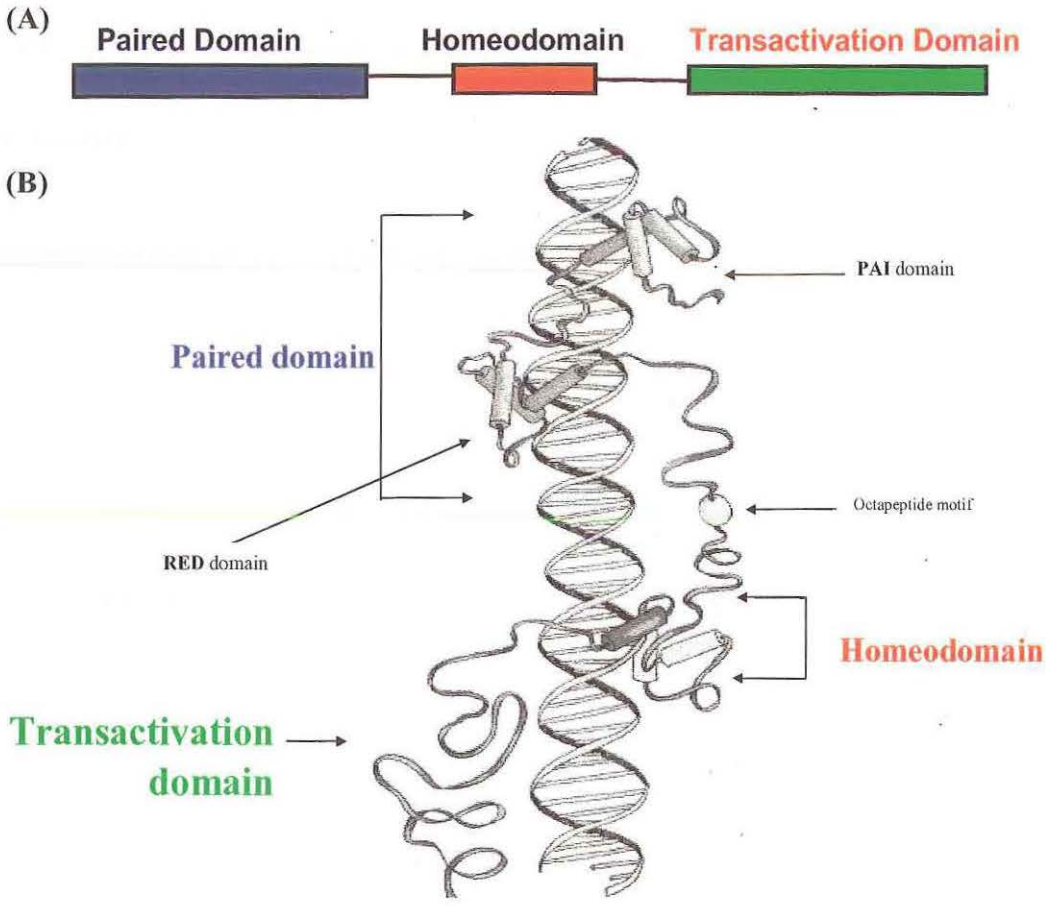


Figure 1.1: Schematic representation of a Pax group III protein (A) showing its two conserved DNA binding domains, Paired domain and homeodomain, and the transactivation domain, and (B) Pax group III protein binding to DNA showing the bipartite paired domain with its PAI and RED domains, the octapeptide, the homeodomain, and the transactivation domain. Barrels in Pax protein represent α-helices.

Four distinct groups of Pax proteins are recognised (Balczarek *et al.*, 1997). Groups III and IV possess both complete PD and HD, group I has only the PD, and group II has a complete PD but only a partial HD (Figure 1.2). It has been proposed that all four groups of *Pax* genes have arisen from a single common ancestor containing both PD and HD DNA binding domains (Balczarek *et al.*, 1997). This gene duplicated very early in Metazoan history giving rise to the progenitors of *Pax* Supergroups I and II, both represented in Cnidaria (Sun *et al.*, 1997). Supergroup I gave rise to the progenitors of *Pax* groups II and IV (Sun *et al.*, 1997). As both of these groups are represented in Cnidaria this duplication predates the Cnidaria-higher Metazoa split (Miller et al., 2000). Supergroup II gave rise to the progenitors of *Pax* groups I and III (Sun *et al.*, 1997). As a group I representative has not been isolated from Cnidaria, it is possible that this duplication occurred after the Cnidaria-higher Metazoa split. However, the *PaxD* gene from the Anthozoan Cnidarian *Acropora millepora* is unequivocally a *Pax* group III gene (Miller et al., 2000), so it may simply be that a group I gene has not yet been isolated within Cnidaria. All four groups are represented in both vertebrates (deuterostomes) and arthropods (protostomes) thus a single gene of each group must have been present in the last common protostome-deuterostome ancestor (Balczarek *et al.*, 1997; Noll, 1993).

Higher vertebrates have nine *Pax* genes (*Pax1-9*) that have arisen via gene duplication from four *Pax* genes present at the divergence of protostomes and deuterostomes (Balczarek *et al.*, 1997). The available data indicate that these duplications have occurred very early in the evolution of vertebrates, as the invertebrate chordate amphioxus (*Branchiostoma floridae*; Cephalochordata) has only single genes of each of the four groups of *Pax* genes (Kozmik *et al.*, 1999; Holland *et al.*, 1999; Holland *et al.*, 1995). Eight *Pax* genes have been identified in *Drosophila melanogaster*, plus a pseudo-*Pax* gene, *Eyegone*, which has a severely deleted PD (Jun *et al.*, 1998). *Pax* genes and their groupings (Balczarek *et al.*, 1997) are displayed in Table 1.1 and their inferred lineage (Balczarek *et al.*, 1997; Sun *et al.*, 1997) is displayed in Figure 1.2.

Table 1.1: *Pax* genes identified in amphioxus, vertebrates and *D. melanogaster*, and their respective groupings defined by Balczarek *et al.* (1997). Genes defined by Groger *et al.* (2000), Holland *et al.* (1999), Kozmik *et al.* (1999), Miller *et al.* (2000), Balczarek *et al.* (1997), Sun *et al.* (1997), and Holland *et al.* (1995).

Pax Group	Amphioxus	Vertebrates	*D. melanogaster*
I	*Amphipax1*	*Pax1*	*Poxmeso*
		Pax9	
II	*Amphipax2/5/8*	*Pax2*	*Sparkling*
		Pax5	
		Pax8	
III	*Amphipax3/7*	*Pax3*	*Paired (Prd)*
		Pax7	*Gooseberry (Gsb)*
			Gooseberry-neuro (Gsb-n)
IV	*Amphipax6*	*Pax4*	*Poxneuro*
		Pax6	*Eyeless*
			Twin of Eyeless (TOY)

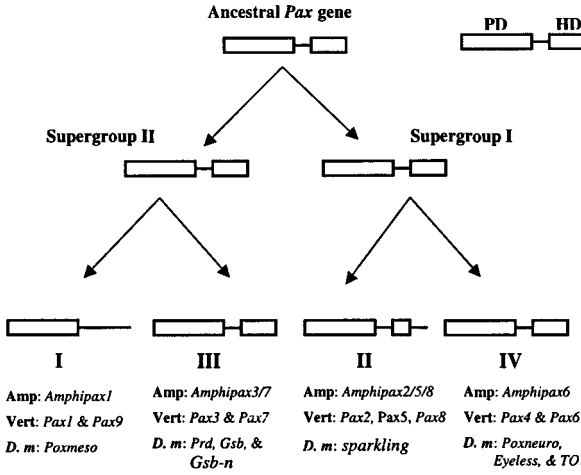


Figure 1.2: The evolutionary lineage of *Pax* genes developed by Sun *et al.* (1997). Groups I, II, III, and IV are those originally proposed by Balczarek *et al.* (1997), supergroups I and II are those originally proposed by Sun *et al.* (1997). Shown are the paired domain (PD) and the homeodomain (HD), the complete or partial deletion of the HD is represented in groups I and II, respectively. Grouping of specific genes; Amp (amphioxus), Vert (vertebrates), *D. m* (*Drosophila melanogaster*) are those proposed by Balczarek *et al.* (1997), Groger *et al.* (2000), and Sun *et al.* (2002).

1.1.1 *Pax* Group III

The deuterostome lineage of *Pax* group III (*pgIII*) genes is relatively well defined. A single *pgIII* gene, *Amphipax3/7*, is found in the cephalochordate *B. floridae* (amphioxus) (Holland *et al.*, 1999), and two *pgIII* genes, *Pax3* and *Pax7*, are found in vertebrates (Seo *et al.*, 1998). This indicates that the *pgIII* gene inherited from the last common protostome-deuterostome ancestor was still present in the earliest chordates, and that this gene duplicated prior to the appearance of vertebrates (Holland *et al.*, 1999).

The protostome *pgIII* lineage is less clearly defined. A single *pgIII* gene has been found in the basal arthropod class, Chelicerata (the mite *Tetranychus urticae*) (Dearden *et al.*, 2002),

and up to three *pgIII* genes are found in the apical arthropod class, Hexapoda (insects) (Baumgartner *et al.*, 1987). Within the insects, two *pgIII* genes are found in the more basal Orthoptera (the grasshopper *Schistocerca americana*) (Davis *et al.*, 2001) and three *pgIII* genes are found in the more apical Diptera (the fruit fly *Drosophila melanogaster*) (Baumgartner *et al.*, 1987). Whereas this may appear as a simple additive progression, from one to two to three *pgIII* genes, the two Orthopteran *pgIII* genes are more closely related to each other than to any of the three Dipteran *pgIII* genes, leading to the assumption that all five genes were derived from a single common ancestor (Davis *et al.*, 2001).

Pax genes have not yet been isolated from taxa evolutionarily positioned between the insects and chelicerates. Crustacea are positioned between the Chelicerata and Hexapoda in the arthropod evolutionary lineage, forming a sister group with the insects (Turbeville *et al.*, 1991). As the positioning of the *pgIII* duplication events leading to insects is uncertain (Davis *et al.*, 2001), it is not known how many *pgIII* genes were present in the last common insect-crustacean ancestor.

1.2 The Functions of *pgIII* Genes

All *pgIII* genes are multifunctional and are involved in the tissue-specific regulation of cell proliferation and differentiation, and through this the specification of body tissue patterning and development of the central nervous system (CNS) (Xue & Noll, 2002). The function of *pgIII* genes in CNS patterning appears to be ancestral, as *pgIII* genes are involved in CNS specification in both *D. melanogaster* and vertebrates (Gutjahr *et al.*, 1993; Maroto *et al.*, 1997).

D. melanogaster has three *pgIII* genes, *Prd*, *Gsb*, and *Gsb-n*. *Prd* and *Gsb* play early roles in larval development and segmentation (Xue *et al.*, 2001). Both *Prd* and *Gsb* are initially expressed in the ectoderm, *Prd* specifies the position of segments along the anteroposterior axis of the developing embryo and *Gsb* specifies polarity of these segments (Baumgartner *et al.*, 1987; Xue & Noll, 2002). *Gsb* also plays an important role in specifying certain developing neurones in the CNS. *Gsb-n* is not involved in embryonic segmentation, but is exclusively expressed in the neuroectoderm (Baumgartner *et al.*, 1987). It is not known

whether any of the *D. melanogaster* genes play a role in myogenesis. However, during embryogenesis both *Prd* and *Gsb* are expressed in the mesoderm (Baumgartner *et al.*, 1987). Other *pgIII* genes isolated from grasshopper, mite and amphioxus have not been functionally studied (Dearden *et al.*, 2002; Davis *et al.*, 2001; Holland *et al.*, 1999).

Vertebrate *Pax3* and *Pax7* are required for development of the CNS and are also important for the regulation of myogenesis (Seale *et al.*, 2000; Goulding, *et al.*, 1994; Jostes *et al.*, 1991). These functions appear to be conserved in the *Pax3* and *Pax7* genes of all vertebrates (mice, humans, birds, and fish) (Seo *et al.*, 1998). Also, similar expression patterns of the nearest common ancestral gene to both *Pax3* and *Pax7*, *Amphipax3/7* (in amphioxus), indicate that these CNS and muscle patterning functions are conserved in early chordates (Holland *et al.*, 1999). In mice, *Pax3* and *Pax7* have been demonstrated to specify embryonic myoblasts and muscle satellite cells, respectively (Seale *et al.*, 2000; Maroto *et al.*, 1997). A function for *pgIII* genes in myogenesis, however, has not been documented outside of the phylum Chordata (*Pax3*, *Pax7* and *Amphipax3/7*).

1.3 Myogenesis

1.3.1 Inductive Signalling and Mesoderm Patterning

In vertebrates, all limb and trunk striated (skeletal) muscle myoblasts arise from the somites of the developing embryo (Francis-West *et al.*, 2003; Christ & Ordahl, 1995). Vertebrate embryonic myogenic determination within the somite is regulated by Wingless-ints (Wnt), Sonic hedgehog (Shh) and Bone morphogenic protein (Bmp) signalling molecules from surrounding tissues (Borycki *et al.*, 1999; Kos *et al.*, 1998). Wnt and Shh signals are both inductive of myogenesis. Wnt peptides (Wnt1 or Wnt3) are released from the neural tube, Shh protein is derived from the floor plate/notochord (Francis-West *et al.*, 2003). These two molecules are sufficient to induce myogenesis in cultured somites (Munsterberg *et al.*, 1995). In contrast, Bmp is inhibitory of muscle differentiation. Bmp is released from the lateral plate mesoderm, and has been shown to block *MyoD* (a myogenic marker) expression (Pourquie *et al.*, 1996). Thus, myogenic patterning in the developing somite is a consequence of Wnt/Shh/Bmp gradients.

1.3.2 The Role of *Pax3* and *Pax7* in Vertebrate Myogenesis

In mice and other vertebrates, *Pax3* is required for the specification of embryonic muscle precursors (embryonic myoblasts), whereas *Pax7* is required for the specification of adult muscle precursors (satellite cells) (Seale *et al.*, 2000; Goulding *et al.*, 1994). *Pax3* and *Pax7* are expressed in the dermomyotome, the *Pax3* expressing cells become embryonic myoblasts, whereas *Pax7* expressing cells become satellite cells (Zammit & Beauchamp, 2001; Bober *et al.*, 1994). Consistent with this idea, *Pax7* null mice have normal embryonic muscle development, but do not develop satellite cells and exhibit no post-natal myogenesis (Seale *et al.*, 2000). Mice with a null mutation in *Pax3* (splotch homozygotes) die *in utero* and exhibit a total absence of limb musculature (Goulding *et al.*, 1994). Aberrant transcripts or chromosomal translocations involving either human *PAX3* or *PAX7* genes have both been identified as causes of rhabdomyosarcoma (skeletal muscle-originating malignant tumours caused by a rampant proliferation of myoblasts) (Tiffin *et al.*, 2003; Barr *et al.*, 1999). This indicates that both *Pax3* and *Pax7* are involved in the regulation of myoblasts, and that certain downstream targets are conserved between the two paralogues (Tiffin *et al.*, 2003; Barr *et al.*, 1999).

Functional studies in P19 carcinoma cells indicate that *Pax3* regulates *Six1*, *Eya2* and *Dach2*, which in turn activate the expression of *MyoD* and *myogenin*, and that *Pax3* is itself upregulated by Wnt3 (reviewed in Parker *et al.*, 2003). A combination of Wnt and Shh signals is sufficient to induce myogenic determination in somitic tissue *in vitro* (Munsterberg *et al.*, 1995). Ectopic expression of *Pax3* in paraxial and lateral plate mesoderm is sufficient to induce expression of *MyoD*, *Myf5* and *myogenin*, without Shh or Wnt signals (Maroto *et al.*, 1997). Wnt and Shh activate, upregulate and maintain both *Pax3* and *Pax7* expression in presegmented and paraxial mesoderm explants (Maroto *et al.*, 1997).

Fitting with the paradigm of regeneration being a recapitulation of embryogenesis, it has been proposed that *Pax7* also regulates *Six1*, *Eya2* and *Dach2* within satellite cells, and that Bmp blocks differentiation of satellite cells/myogenic progenitor cells and Shh and/or Wnt induce myogenic differentiation (Parker *et al.*, 2003). To this end it has recently been demonstrated that Wnt is sufficient to induce adult muscle-derived $CD45^+$ stem cells

(muscle side population [SP] cells) to undergo myogenic specification during muscle regeneration (Seale *et al.*, 2003) and that injection of a Wnt antagonist (sFRP) into regenerating muscle markedly reduced proliferation and myogenic specification of $CD45^+$ SP cells (Polesskaya *et al.*, 2003). Also, treating cultured satellite cell-derived myoblasts or satellite cells (on muscle fibres) with Bmp results in their differentiation into osteocytes, occurring concurrently with the suppression of *Pax7* and *MyoD* (Asakura *et al.*, 2001).

In *Pax7* null mice, muscle SP stem cell numbers are unaltered but satellite cells are absent and post-natal myogenesis does not occur (Seale *et al.*, 2000). Asakura *et al.* (2002) demonstrated that, following intra-muscular injection of muscle-derived stem cells (which do not express *Pax7*), these cells readily differentiated into satellite cells coupled with the concomitant expression of *Pax7*. Taken together these data strongly imply a role for *Pax7* in the myogenic specification of adult stem cells.

Expression of both *Pax3* and *Pax7* is induced by Wnt signals, and both *pgIII* genes are upstream regulators of the expression of the *MyoD* family of myogenic regulatory factor genes and *Mef2* in myogenesis (Polesskaya *et al.*, 2003; Tajbakhsh, 2003; Schmidt *et al.*, 2000; Borycki *et al.*, 1999; Maroto *et al.*, 1997). This indicates an important role for vertebrate *pgIII* genes in mediating the myogenic inductive effects of Wnt/Shh/Bmp pathways and the myogenic determination and differentiation MyoD/Mef2 pathways. This *pgIII* function is myoblast specification.

1.3.3 Vertebrate Striated Muscle Determination and Differentiation

Two pathways are required for the differentiation of striated muscle; the MyoD family pathway and the Mef2 pathway (Dodou *et al.*, 2003; Sabourin & Rudnicki, 2000). Members of the MyoD family are responsible for the determination and differentiation of both embryonic and adult myoblasts into striated muscle (Sabourin & Rudnicki, 2000; Robertson, 1990). Higher vertebrates have four *MyoD* family genes that operate in a functional hierarchy. *MyoD* and *Myf5* are responsible for myogenic specification, and *MRF4* and *myogenin* act downstream of *MyoD/Myf5* in myogenesis, and are responsible for terminal differentiation (Sabourin & Rudnicki, 2000). Together, these members of the

MyoD family are responsible for the proliferation and differentiation required for the conversion of myoblasts into multinucleated muscle cells (Sabourin & Rudnicki, 2000).

The role of *MyoD* family in embryonic and adult myogenesis is highly analogous (Tajbakhsh, 2003; Asakura *et al.*, 2001). In *Pax3* expressing myoblasts of the dermomyotome, *Myf5*, *MyoD*, *myogenin* and *MRF4* become activated and direct myogenic differentiation (Sabourin & Rudnicki, 2000). Quiescent *Pax7*-specified satellite cells do not express any *MyoD* family genes, however once activated either *MyoD* or *Myf5* are rapidly upregulated, followed by their co-expression directing satellite cell proliferation (Sabourin & Rudnicki, 2000). After several rounds of proliferation, *myogenin* and *MRF4* are expressed, directing terminal differentiation (Sabourin & Rudnicki, 2000).

Mef2 factors are essential for muscle differentiation; the ability of the MyoD family transcription factors to convert cells into striated muscle is dependent upon the function of Mef2 factors. *Mef2* is activated by MyoD and regulates expression of *myogenin* (Dodou *et al.*, 2003). Mef2 regulates genes involved in the fusion and terminal differentiation of muscle cells. Mice with an inactivated *Mef2c* gene die *in utero* due to incomplete activation of downstream striated muscle structural genes, but do not show any defects in myoblast specification (reviewed in Dodou *et al.*, 2003).

1.3.4 Myogenic Pathways in Invertebrates

The formation of muscle during *D. melanogaster* embryogenesis utilises many similar genetic pathways (Baylies & Michelson, 2001: Michelson, 1994). The Wnt/Shh/Bmp inductive signalling pathways and the Mef2/MyoD myogenic pathways are also found in *D, melanogaster. D. melanogaster* has direct structural and functional homologues of Wnt (Wg; wingless), Shh (Hh; hedgehog) and Bmp (Dpp; decapentaplegic) Mef2 and MyoD (Nau; Nautilus) (Dodou *et al.*, 2003; Akam, 2000; Morata & Sanchez-Herrero, 1999; Lin *et al.*, 1997; Ranganayakulu *et al.*, 1996).

Not only are these pathways highly conserved between *D. melanogaster* and vertebrates, Wnt/Bmp and MyoD/Mef2 pathways are also present in Cnidaria (Muller *et al.*, 2003; Schmidt *et al.*, 2000). *Wingless* (*Wnt*) homologues have been isolated in Crustacea, but

have not been studied in relation to myogenesis (Williams & Nagy, 2001; Nulsen & Nagy, 1999).

A great deal less is known about embryonic muscle formation in Crustacea. Although freshwater crayfish have direct development (they do not have free-living larval stages), their embryogenesis occurs in two distinct steps (Scholtz, 2000; Hamr, 1992). Firstly the nauplius is formed (which is free-living in some species, but in *Cherax destructor* is contained within the egg). The nauplius has three segments (corresponding to later head segments), no appendages, but does contain some trunk muscles (Scholtz & Dohle, 1996). After the nauplius stage, the rest of the adult body plan develops (Scholtz & Kawai, 2002). The caudal papilla elongates, segments, and then limb primordia begin to emerge from it. Freshwater crayfish segment twice, first in the development of the nauplius, then second after the extension of the post-naupliar germ band (Scholtz & Kawai, 2002; Scholtz & Dohle, 1996; Scholtz, 1995; Sandeman & Sandeman, 1991). Embryonic development, therefore, is different in Crustacea, than in either vertebrates or insects such as *D. melanogaster*.

1.4 Muscle Regeneration

Striated muscle, in both crustaceans and vertebrates, is a highly adapted tissue specifically constructed for the generation of force by contraction (Li & Mykles, 1990; Bagshaw, 1993). The tissue is composed of aligned syncytial (multinucleate) myofibres whose hundreds of post-mitotic nuclei generate and maintain the contractile and regulatory proteins responsible for regulated force generation. This high degree of specialisation is achieved by terminal differentiation of muscle cell nuclei, which involves the expression of a particular complement of tissue-specific genes accompanied by withdrawal from the cell cycle (Zammit & Beauchamp, 2001). As such, terminally differentiated muscle cells are incapable of self-renewal, and must rely on the contribution made by other cells for growth and repair. The regeneration of striated muscle is thus a process whereby undifferentiated progenitor cells become committed to the myogenic program and fuse to a new or existing myofibre (Grounds & Yablonka-Reuveni, 1993). These progenitor cells undergo the stepwise process of myogenic determination, followed by proliferation (and often migration) and finally terminal differentiation into muscle (Zammit & Beauchamp, 2001).

Muscle satellite cells are the main progenitor cells that fulfil this role, and clearly contribute the majority of post-natal myogenic nuclei in vertebrates (Asakura *et al.*, 2001; Zammit & Beauchamp, 2001; Sabourin & Rudnicki, 2000) and Crustacea alike (Govind & Pearce, 1994). Satellite cells reside closely juxtaposed to, and within the basal lamina of, each myofibre (Grounds *et al.*, 2002). In normal post-natal muscle, satellite cells are mitotically quiescent. However, when supplied with the required signals (in response to weight-bearing exercise, injury, etc.) satellite cells become activated and re-enter the cell cycle. The daughter cells of satellite cells (termed myogenic precursor cells) then undergo several rounds of division followed by fusion to a myofibre (Zammit & Beauchamp, 2001).

In addition to satellite cells (which are muscle progenitor cells, not true stem cells; Asakura *et al.*, 2002), stem cells including circulating haematopoietic stem cells (HSCs) are considered to act as myoblast progenitors in vertebrates (LaBarge & Blau, 2002; reviewed in Grounds *et al.*, 2002) and crustaceans (Erri Babu, 1987; Read & Govind, 1998; Uhrik *et al.*, 1989). In vertebrates (mice) HSCs have been shown to contribute to regenerating striated muscle *in vivo* (LaBarge & Blau, 2002; Gussoni *et al.*, 1999) and also to be transformed into functional myoblasts *in vitro* (Ferrari *et al.*, 2001; Gussoni *et al.*, 1999).

1.4.1 The Regeneration of Striated Muscle in *Cherax destructor*

The yabby *Cherax destructor* (and other decapod crustaceans) possesses a phenomenal capability for almost limitless adult myogenesis. *C. destructor* regenerates muscle as an adult in two distinct ways. Firstly, they grow as adults (unlike some other invertebrates) and so their muscle also grows, and secondly, they are able to regenerate entire limbs (including the muscle within them). The interesting thing about these dichotomous muscle regeneration strategies is that they utilise different stem cells as myoblasts. Endogenous satellite cells provide the majority of nuclei for normal muscle growth (Govind & Pearce, 1994), but in limb regeneration, myoblasts must come from beyond the myofibre.

As crustaceans have a growth-restricting exoskeleton, they cannot simply grow muscle in a linear fashion, muscle growth must be cyclical. Normal growth is strictly regulated around the moult cycle, providing a long period of myogenic quiescence (intermoult) followed by a relatively short period of frenzied myogenic activity coinciding with the shedding of their

exoskeleton (premoult, ecdysis and postmoult) (Lamey *et al.*, 2002; West, 1997; Mykles & Skinner, 1982). This moult-cycle myogenic activity is exacerbated in the claw closer muscle of *C. destructor* (Lamey *et al.*, 2002; West, 1997), which must be withdrawn through the narrow basi-ischial joint at the proximal end of the cheliped limb at ecdysis, a tricky procedure that can sometimes result in the loss of the limb (personal observation). The claw closer muscle atrophies by as much as 60 percent in premoult, and then rapidly regenerates (and grows) following ecdysis (Lamey *et al.*, 2002; Mykles & Skinner, 1982). The regeneration of atrophied muscle begins during the premoult phase (West *et al.*, 1995; Holland & Skinner, 1976).

Muscle satellite cells play a vital role in the moult-cycle regeneration process of crustaceans, contributing the majority of nuclei to the regeneration process (Govind & Pearce, 1994; Uhrik *et al.*, 1989). Satellite cells are quiescent throughout intermoult periods. In response to moulting cues (ecdysteroids), satellite cells are activated, they then proliferate, differentiate and fuse to form multinucleate myofibres, and through this result in the regeneration of muscle (Hopkins *et al.*, 1999; Uhrik *et al.*, 1989). Morphological studies have detailed a similar conversion of satellite cells into myoblasts during muscle regeneration in the snapping shrimp, *Alpheus heterochelis* (Govind & Pearce, 1994) to that observed in higher vertebrates (reviewed in Grounds & Yablonka-Reuveni, 1993). Decapod crustaceans have a several-fold greater density of satellite cells than do vertebrates (Uhrik *et al.*, 1989). In addition, ultrastructural studies indicate that crustaceans utilise haemocytes (haematopoietic stem cells) in the regeneration of myofibres after moulting (Govind & Pearce, 1994; Uhrik *et al.*, 1989). In mechanically damaged muscle of the crayfish *Astacus fluviatilis*, haemocytes have been observed to invade the area, send out cytological projections to endogenous cells (probably satellite cells), then transform into cells resembling myotubes containing contractile material (Uhrik *et al.*, 1989).

1.4.2 Limb Regeneration in Crustaceans

Limb regeneration in crustaceans is epimorphic, in that a blastema is formed at the site of limb loss, which grows and differentiates into the entire limb (Read & Govind, 1998). Crustaceans are adept at epimorphic limb regeneration, and often prefer to autotomise and regenerate an entire limb, rather than repair a damaged limb (Uhrik *et al.*, 1989).

Endogenous satellite cells cannot participate in this process, as all of the muscle is removed along with the limb. Studies in other decapod crustaceans indicate that myoblasts utilised in this muscle regeneration process are derived from immigrant haematopoietic stem cells (Hopkins *et al.*, 1999; Erri-Babu, 1987; Kao & Chang, 1997; Read & Govind, 1998; Uhrik *et al.*, 1989). This research describes a progression beginning with a layer of proliferating epithelial cells, followed by the provision of a structural framework by fibroblasts and the inundation of the whole structure with immigrant haemocytes (Read & Govind, 1998). Haemocytes have been observed to be involved in phagocytosis, protein recycling, and as a source of myogenic stem cells in this process (Read & Govind, 1998). Multiple studies on crustaceans indicate that these immigrant haemocytes transform into both blastema cells and myoblasts during epimorphic regeneration (Read & Govind, 1998; Pearce *et al.*, 1997; Uhrik *et al.*, 1989; Erri Babu, 1987).

Epimorphic regeneration in *C. destructor* occurs in defined stages (Cutler *et al.*, 2002); after limb autotomy the wound seals creating a blastema, which erupts forming a papilla that eventually differentiates and segments (the closed dactyl stage) and, following ecdysis, emerges as a small fully-functional replica of the original limb (Cutler *et al.*, 2002). The papilla stage is characterised by the presence of masses of undifferentiated cells, but once the closed dactyl stage is reached small differentiated myofibres expressing their characteristic protein assemblage can be discerned (Cutler *et al.*, 2002), indicating that the myogenic determination of immigrant stem cells occurs during the papilla stage (Cutler *et al.*, 2002.

The molecular mechanisms behind this phenomenal regenerative capacity are completely unknown. Therefore, in any attempt to elucidate the molecular apparatus that facilitates this process, comparisons must be drawn from other animals for which molecular mechanisms behind similar processes have been uncovered. These animals are the mouse (and to some extent humans) and the fruit fly.

1.4.3 Vertebrate Muscle Regeneration

Although meagre and limited by comparison with crustaceans, the striated (skeletal) muscle of vertebrates actually possesses a remarkable ability to regenerate throughout the lifespan of the animal. For quite some time, the accepted doctrine has been that muscle satellite cells are entirely responsible for post-natal muscle regeneration in higher vertebrates (Mauro, 1961). Satellite cells are clearly the dominant myogenic precursor cell responsible for muscle regeneration, as in crustaceans (Goldring *et al.*, 2002; Grounds *et al.*, 2002; Hawke & Garry, 2001; Zammit & Beauchamp, 2001; Bischoff, 1990). In response to injury, satellite cells become activated and begin to proliferate, migrate to the site of regeneration, terminally differentiate and fuse to the myofibre (or fuse together to form a new myofibre) (Zammit & Beauchamp, 2001). Vertebrate satellite cells are thought to be self-renewing, asymmetrical division allows some activated satellite cell progenitors to form myoblasts while others regain the dormant satellite cell niche (reviewed in Parker *et al.*, 2003). It was the historical view that this was the only way that adult myogenesis was effected. Recent experiments, however, have expanded this view quite considerably (Asakura *et al.*, 2002; LaBarge & Blau, 2002; Jankowski *et al.*, 2002; Qu-Petersen *et al.*, 2002; Gussoni *et al.*, 1999).

Satellite cells have long been characterised exclusively on morphological criteria (Seale & Rudnicki, 2000). Recent attempts to extract and characterise satellite cells on biochemical and immunological criteria have revealed multipotent stem cells (distinct from satellite cells) are present in adult striated muscle (Asukura *et al.*, 2002; Asakura *et al.*, 2001). We now know that vertebrate muscle is filled with a heterogeneous compartment of stem cells (Qu-Petersen *et al.*, 2002). These cells, termed muscle side population (SP) cells, were first identified by their fluorescence-activated cell sorting (FACS) exclusion of Hoechst dye (Gussoni *et al.*, 1999; Asakura *et al.*, 2001; Asakura *et al.*, 2002).

Immunological criteria that have been useful in characterising these cells are cell surface proteins CD45 and Sca1 (both expressed on haematopoietic stem cell lineages) (McKinney-Freeman *et al.*, 2002; Jankowski *et al.*, 2002). Muscle SP cells express these markers; satellite cells do not express either marker ($CD45^-/Sca1^-$) (Asakura *et al.*, 2002). Moreover, these adult muscle-derived SP cells are somewhat developmentally plastic and are capable

of repopulating many other lineages in lethally irradiated mice (haematopoietic, osteogenic, lipogenic), whereas satellite cells do not appear to have this capability and are committed to the myogenic lineage (McKinney-Freeman *et al.*, 2002; Asakura *et al.*, 2002).

Muscle SP cells participate in muscle regeneration (Polesskaya *et al.*, 2003; Seale *et al.*, 2003), and the first step in their myogenic conversion is to occupy the satellite cell niche (Asakura *et al.*, 2002). $CD45^+$ SP cells become satellite cells and undergo myogenesis when injected into regenerating muscle, when cocultured with regenerating myoblasts, or when supplied with the necessary inductive signals (Asakura *et al.*, 2002; Polesskaya *et al.*, 2003). Interestingly, the origin of $CD45^+$ muscle SP cells is unknown. It is possible that they are derived from the somites of the developing embryo or that they are replenished systemically from bone marrow or vascular sources (Snider & Tapscott, 2003; McKinney-Freeman *et al.*, 2002; Tamaki *et al.*, 2002), or even a combination of all three sources is not out of the question.

Adult stem cells from elsewhere in the body can participate in myogenesis in vertebrates (Ferrari *et al.*, 1998). Stem cells derived from a wide variety of adult tissues are capable of becoming muscle (reviewed in Goldring *et al.*, 2002 and Grounds *et al.*, 2002), however the occurrence of this is so infrequent that its biological relevance is almost negligible (Grounds *et al.*, 2002). LaBarge and Blau (2002) have demonstrated that, following bone marrow transplantation, adult bone marrow-derived stem cells participate in myogenesis and become muscle satellite cells. This process occurred markedly following muscle irradiation (which destroys endogenous satellite and SP cells) and also in response to exercise-induced muscle damage (contributing to 3.5% of muscle fibres; LaBarge & Blau, 2002). This demonstrates that circulating haematopoietic stem cells may participate in muscle regeneration under extreme (yet still biologically relevant) circumstances in adult mice, and they do this by first converting to satellite cells (LaBarge & Blau, 2002).

In summary, satellite cells perform adult muscle regeneration. Satellite cells are a self-renewing lineage derived embryonically from the somites, however, other non-satellite stem cells can replenish the satellite cell compartment (and thus participate in muscle regeneration) in times of need. These non-satellite stem cells are predominantly the muscle SP compartment (resident within muscle) but can also be cells derived from other sources,

specifically haematopoietic stem cells. As in crustaceans, muscle satellite cells are the dominant source of myogenic precursor cell, but under extreme circumstances haematopoietic stem cells may also participate in this role.

1.4.4 The Role of *Pax7* in Muscle Regeneration

Pax7 occupies a pivotal role in the regeneration of adult striated muscle; it specifies satellite cells (Seale *et al.*, 2000). Satellite cells may be derived from the somites of the developing embryo, from muscle SP cells, or from haematopoietic stem cells (Polesskaya *et al.*, 2003; Asakura *et al.*, 2002; LaBarge & Blau; 2002). *Pax7* is expressed in both quiescent and activated satellite cells and is downregulated during terminal differentiation (Seale *et al.*, 2000). Mice that lack *Pax7* do not have any satellite cells and do not undergo any postnatal muscle growth, and as a consequence die two weeks after birth (Seale *et al.*, 2000).

Striated muscle from *Pax7* null mice does not contain satellite cells, but contains normal numbers of SP cells that display an increased ability to differentiate along the haematopoietic lineage *in vitro* (Seale *et al.*, 2000; Asakura *et al.*, 2002), indicating that the induction of *Pax7* converts SP cells into myoblasts/satellite cells by restricting other developmental lineages. The induction of *Pax7* through extracellular inductive signals (Wnt) has been demonstrated to specify the conversion of $CD45^{+}$ SP cells into satellite cells, and facilitates their myogenic determination; and blocking of these inductive signals precludes the conversion of SP cells to satellite cells (Polesskaya *et al.*, 2003).

The well-established role for vertebrate *pgIII* genes *Pax3* and *Pax7* in the specification of embryonic myoblasts and muscle satellite cells, respectively, and the requirement of *Pax7* for adult muscle regeneration (Seale *et al.*, 2000), singled out *pgIII* genes as appropriate candidates to research in adult muscle regeneration in *C. destructor*. The aim of this research is to identify *pgIII* genes in *C. destructor* and determine whether these genes are expressed during adult muscle regeneration in *C. destructor*. The characterisation of *Pax* genes in Crustacea is also of significance to our understanding of the evolution of this important family of transcription factors.

1.5 Research Aims

1.5.1 Main Aim

To identify *Pax* group III genes in *Cherax destructor* and to ascertain if these genes are expressed during adult myogenesis.

1.5.2 Specific Aims

Aim 1: Isolate *Pax* group III gene(s) from *C. destructor*.

1.1 Design degenerate primers based on homologous regions of *Pax* group III genes from evolutionary relatives of *C. destructor*.

1.2 Amplify, isolate and sequence *Pax* group III genes from *C. destructor* genomic DNA.

1.3 Amplify, isolate and sequence *Pax* group III sequences from *C. destructor* mRNA from embryonic tissue.

1.4 Identify gene(s) or cDNAs from sequences.

Aim 2: Characterise the protein(s) of *C. destructor Pax* group III genes.

2.1 Derive protein sequences from DNA/cDNA sequences.

2.2 Determine homology with other Pax group III protein sequences.

2.3 Determine phylogenetic relationships of Pax proteins, and Pax group III proteins in particular.

2.4 Deduce putative functions of the *C. destructor* Pax proteins by comparison to other Pax proteins of known function.

Aim 3: Establish expression pattern of *Pax* group III gene(s) in myogenesis in *C. destructor*.

3.1 Amplify *Pax* group III mRNA from whole embryos during times of myogenesis.

3.2 Amplify *Pax* group III mRNA from regenerating claw closer muscle during normal postnatal growth.

3.3 Amplify *Pax* group III mRNA from epimorphically regenerating cheliped tissue during phases of muscle regeneration.

2.0 METHODS

2.1 Research Plan

The purpose of this research was to isolate *Pax* group III (*pgIII*) genes from the Australian freshwater crayfish *Cherax destructor* (the yabby) and assay *pgIII* gene expression in epimorphic and moult-cycle regenerating muscle. To isolate *pgIII* genes from *C. destructor*, nested partially-degenerate primers for polymerase chain reaction and reverse transcriptase - polymerase chain reaction (PCR and RT-PCR, respectively) were designed. These primers were designed to complement highly conserved regions of *pgIII* genes from close evolutionary relatives of Crustacea (specifically mite, grasshopper and fruit fly). To test the specificity of these primers, and the techniques being used, to *pgIII* genes, it was necessary to ensure that they amplified all of the *pgIII* genes from a positive control without producing non-specific artefacts. Positive controls consisted of DNA (from whole flies) and RNA (from embryos) isolated from *Drosophila melanogaster*, which has three *pgIII* genes.

Primers were used to amplify *pgIII* genes from *C. destructor* DNA using PCR. Due to the likelihood of introns resulting in the amplified region in genomic DNA exceeding the limits of the technique, primers were also used to amplify *pgIII* sequences from RNA isolated from embryos at 30% and 50% of development using RT-PCR. All other *pgIII* genes characterised in arthropod species to date are expressed in embryos at 30% and/or 50% of development (Dearden *et al.*, 2002; Davis *et al.*, 2001; Zhang *et al.*, 1994; Gutjahr *et al.*, 1993; Kilcherr *et al.*, 1986). Amplification products were then gel purified and sequenced directly. Expression of isolated *pgIII* genes in intermoult, premoult, postmoult and epimorphically regenerating muscle from the claw closer muscle (or pre-muscle tissue) of *C. destructor* was assayed by RT-PCR.

Open reading frames and protein sequences were deduced from the obtained *pgIII* cDNA sequences. *C. destructor* pgIII protein sequences were analysed and their phylogenetic relationship to other known Pax proteins was inferred by maximum likelihood analysis.

2.2 Experimental Animals

2.2.1 Animal Care

C. destructor (Crustacea; Decapoda; Parastacidae) were obtained from a local supplier (Aquaculture Solutions) and from a colony kept at Edith Cowan University (ECU). Animals were kept in 200 x 250 x 350mm glass aquaria containing approximately 6cm depth of shell grit substrate and lengths of PVC pipe for cover. Aquaria were half filled with half tap water/half rainwater (changed fortnightly), constantly oxygenated and maintained at a water temperature of 19°C (± 2°C). Animals were fed *ad libitum* on chicken pellets, algae pellets, peas and corn. Animals with an ocular carapace length (OCL; Figure 2.1) between 20 and 30mm were used for muscle and regenerate limb sampling, as small *C. destructor* moult with increased frequency (West, 1997). Larger animals (OCL >35mm) were utilised in breeding for embryo production.

2.2.2 Breeding of *C. destructor*

Female *C. destructor* used for breeding were selected on the basis of size (OCL >35mm). Male *C. destructor* selected for breeding were slightly larger than the females with which they were grouped. Breeding groups consisting of one male and two female animals were placed in aquaria until one of the females became berried. Berried females were immediately isolated and maintained as such throughout embryonic development, taking 40 days in *C. destructor* at 19°C (Sandeman & Sandeman, 1991). Five breeding groups were established and three females became berried over the course of this study.

C. destructor embryos were harvested from non-anaesthetised animals at four day intervals throughout development. Female *C. destructor* carry embryos underneath their tail, and thus the embryos are easily accessible (Sandeman & Sandeman, 1991). Females were carefully manually restricted (to prevent tail flick and resultant abortion) and single embryos were plucked from beneath the tail using forceps. The developmental stages of embryos were assessed by both developmental time and embryo morphology, using the methods of Sandeman and Sandeman (1991). Embryos were whole mounted in 70% ethanol on concave glass microscope slides and viewed under a dissecting microscope.

Embryo stages were photographed using a Kodak digital camera mounted to a dissecting microscope. Embryos at 30% (naupliar) and 50% (post-naupliar) of development were used in this study (Sandeman & Sandeman, 1991). Embryos were immediately snap-frozen in liquid nitrogen and stored in Eppendorf tubes at -80°C until required (<2 months).

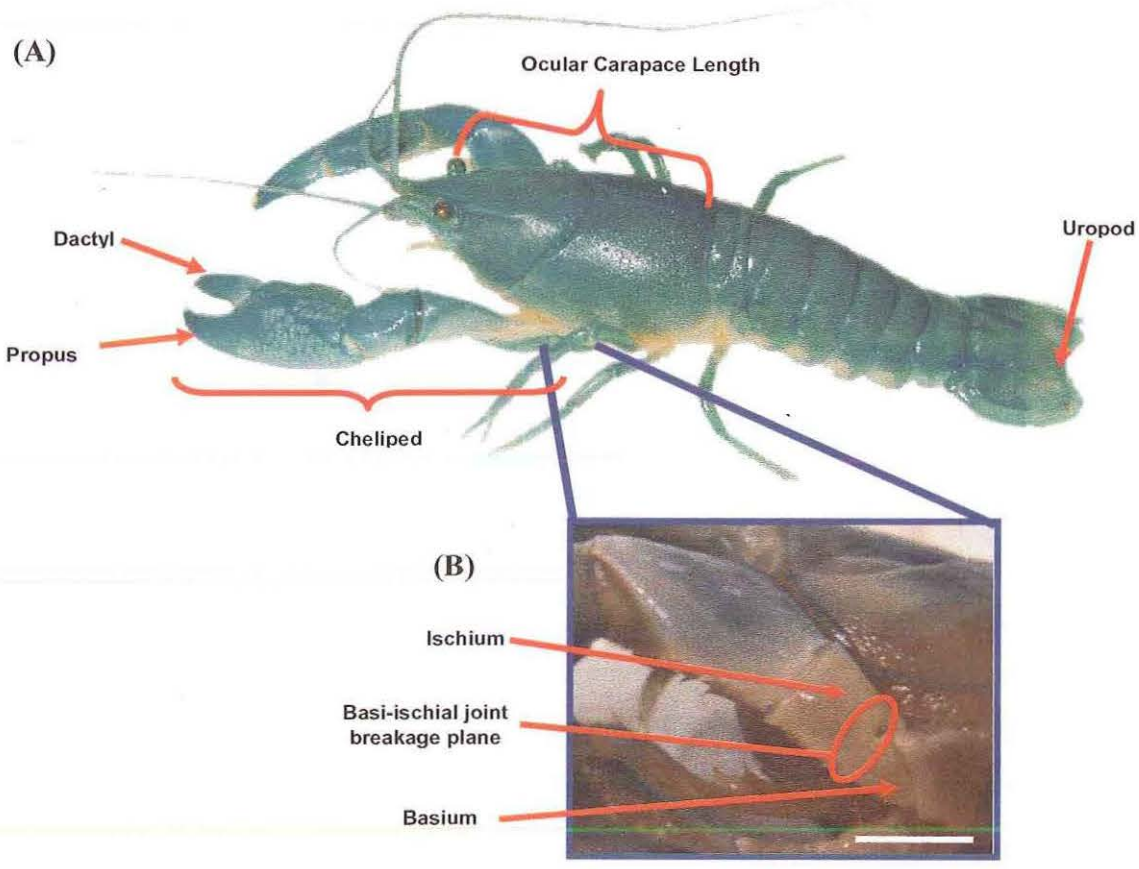


Figure 2.1: Adult *Cherax destructor* (A) showing morphological characters (to scale), and (B) close-up of basi-ischial joint breakage plane (*scale bar*: 10mm).

2.2.3 Sampling of Moult-Cycle Muscle Regeneration in *C. destructor*

Moult cycle stage was determined by uropod setal morphology as described by Burton and Mitchell (1987). Sections of the posterior margin of the uropods (Figure 2.1) were removed and placed ventral-side-up on glass microscope slides, covered with distilled

water and a coverslip, and examined under a compound microscope. The moult cycles of three male and three female *C. destructor* were monitored twice weekly (except when animals were nearing ecdysis, when moult cycles were monitored daily). Claw closer muscle samples were taken at late premoult (stages D_2 to D_4; Burton & Mitchell, 1987), immediately postmoult (stage A; Burton & Mitchell, 1987), and at intermoult (stage C_1 or C_2; Burton & Mitchell, 1987). Cheliped autotomy was performed by removing an entire cheliped (either left or right) from ice anaesthetised animals by snapping the cheliped at the basi-ischial joint breakage plane (Figure 2.1) using forceps. Muscle samples were taken from autotomised chelipeds by opening the claw carapace dorsally and dissecting out a $5mm^3$ section of the proximal section of the claw closer muscle. Autotomised chelipeds were kept on ice throughout the procedure. Muscle samples were immediately snap-frozen in liquid nitrogen and stored in Eppendorf tubes at -80°C until required for RNA extraction (<2 months).

2.2.4 Sampling of Epimorphic Regeneration in *C. destructor*

Three male and three female *C. destructor* were induced to autotomise either their left or right cheliped at intermoult by removing the limb at the basi-ischial joint breakage plane (as above) (Figure 2.1: B). The regeneration of autotomised chelae was monitored and entire regenerate limb samples were taken at the undifferentiated papilla and closed dactyl stages (Cutler *et al.*, 2002). To do this, animals were ice anaesthetised and entire regenerate limbs were removed at the basi-ischial joint, immediately snap-frozen in liquid nitrogen and stored in Eppendorf tubes at -80°C until required for RNA extraction (<2 months).

2.2.5 *Drosophila melanogaster*

Adult wild type *Drosophila melanogaster* (fruit fly; Hexapoda; Insecta; Diptera) were obtained from a stock bred at ECU. Flies aged between 4 and 25 days were used for embryo production as at this age *D. melanogaster* are most fertile (Weichaus & Nusslein-Volhard, 1986). Stocks were bred in 25mL clear plastic tubes capped with foam containing breeding medium (45.0g.L^{-1} glucose, 22.5g.L^{-1} sucrose, 11.6g.L^{-1} ethanol, 10.1g.L^{-1} yeast, 6.6g.L^{-1} potassium tartrate, 4.5g.L^{-1} agar, 4.1g.L^{-1} propionic acid, 1.3g.L^{-1} methylparaben, 0.4g.L^{-1} calcium chloride, 0.4g.L^{-1} phosphoric acid) at 22°C. Four days after eclosure, flies

were transferred from breeding tubes to laying chambers for embryo production. Flies in laying chambers were cleared after three weeks to maintain productivity, with approximately 100 flies being maintained in each laying chamber concurrently. Laying chambers were maintained at 25°C during laying periods and at 22°C between laying periods.

2.2.6 *D. melanogaster* Embryo Collection

Laying chambers were constructed from 1L clear plastic jars. Holes (4.5cm diameter) were cut into the side of each jar into which foam was inserted allowing airflow. Jars were upturned onto a Petri dish (laying plate) containing laying medium (24.3g.L^{-1} agar, 21.4g.L^{-1} sucrose, 1.4g.L^{-1} methylparaben, 24%v/v Berri apple juice) laced with a 2cm^2 drop of live yeast suspension (286g.L^{-1} Tandaco live bakers yeast suspended in diH_2O).

Maximal expression of *Prd* occurs at stage 8 of embryonic development, whereas maximal expression of *Gsb* and *Gsb-n* occur at stage 14 (Kilcherr *et al.*, 1986). *D. melanogaster* embryonic development is temperature dependent; at 25°C stage 8 is reached 4 hours after laying and stage 14 is completed 14 hours after laying (Campos-Ortega & Hartenstein, 1985). Timed embryos were obtained by adding fresh laying plates to laying chambers and incubating at 25°C for 14 hours, then removing and incubating plates for a further 4 hours at 25°C. This ensured obtaining a mixture containing both stage 8 and stage 14 *D. melanogaster* embryos, from which *Prd*, *Gsb*, and *Gsb-n* mRNA could be extracted.

Embryos were manually collected from plates under a dissecting microscope and transferred to a glass coverslip using a metal probe. Embryos were then snap-frozen by immersion of the loaded coverslip into liquid nitrogen. Frozen embryos were then transferred to a DEPC-treated 1.7mL Eppendorf tube and stored at -80°C until required (<2 months). Approximately 50 embryos were collected in each sample.

2.3 Nucleic Acid Extraction

2.3.1 Working Conditions

All glassware, metal implements and ceramic ware used for DNA extraction was either baked for 3 hours at 170°C or autoclaved prior to use. Disposable Eppendorf tubes, reaction vessels and filtered pipette tips were used throughout.

All glassware, metal implements and ceramic ware used for RNA extraction was baked at 170°C for at least 3 hours prior to use. All plastic ware was either certified RNase free or treated with diethyl pyrocarbonate (DEPC) and autoclaved prior to use. All water used for dilution of RNA solutions was treated with 0.1%w/v DEPC and autoclaved. Pipette ends were soaked in 2.5M HCl, washed thoroughly with DEPC diH_2O and dried in a UV light box before use.

2.3.2 DNA Extraction from *C. destructor* and *D. melanogaster*

Total genomic DNA was extracted using the Wizard genomic DNA purification kit (Promega) as stipulated by the manufacturer. Approximately 100mg of freshly dissected claw closer muscle tissue (*C. destructor*) or 20 freshly sacrificed whole adult flies (*D. melanogaster*) was ground under liquid nitrogen using a mortar and pestle. Ground tissue was transferred to a 1.7mL Eppendorf tube containing 600μL chilled nuclei lysis solution and incubated at 65°C for 30 minutes. This solution was cooled, 3μL of 4mg.mL^{-1} RNase A solution was added, the tube was mixed by inversion, incubated at 37°C for 30 minutes and allowed to cool to room temperature. Protein precipitation solution (200μL) was added and the tube was vortexed at high speed for 30 seconds, placed on ice for 5 minutes and centrifuged at 13 000g for 4 minutes at room temperature. The supernatant was transferred to a new tube containing 600μL room temperature isopropanol, mixed by inversion and centrifuged at 13 000g for 1 minute. The supernatant was discarded and the resultant pellet was washed with 600μL 70%v/v ethanol and centrifuged at 13 000g for 1 minute at room temperature (twice). All supernatant was discarded, the pellet was air dried for 10 minutes at room temperature and then resuspended in 100μL of TE buffer (10mM Tris-HCl pH 8, 1mM EDTA) by incubation at 65°C for 1 hour with periodic mixing by inversion.

2.3.3 RNA Extraction from *D. melanogaster* Embryos

Chilled (4°C) Trizol reagent (Invitrogen) (1mL) was added to thawed *D. melanogaster* embryos in 1.7mL Eppendorf tubes. The resulting suspension was vortexed at high speed and transferred to a glass micro-mortar and pestle on ice. The embryos were homogenised, assessed under a dissecting microscope and transferred to a new DEPC-treated tube. The solution was then drawn through a 21 gauge needle four times (to dissociate RNA from DNA) and stored at 4°C for 5 minutes (to allow complete dissociation of nuclear-protein complexes). Chloroform (200μL) was added with 30 seconds of vigorous manual agitation, the solution was then stored at 4°C for 5 minutes and centrifuged at 12 000g for 15 minutes at 4°C. The supernatant was transferred to a new DEPC-treated tube, an equal volume of room temperature isopropanol was added and the tube was stored at -20°C for 30 minutes, centrifuged at 12 000g for 10 minutes at 4°C and the supernatant discarded. The resultant pellet was washed with 70%v/v ethanol (in DEPC diH_2O) and centrifuged at 7 500g for 5 minutes at 4°C (twice). The washed pellet was air dried on ice for 10 minutes, resuspended in 40μL DEPC diH_2O with vigorous vortexing and stored at -80°C until required (<3 months).

2.3.4 RNA Extraction from *C. destructor* Tissue and Embryos

Approximately 100mg muscle tissue, regenerate limb tissue, or 10 embryos were homogenised in a mortar and pestle under liquid nitrogen. Homogenate was transferred to a DEPC-treated 1.7mL Eppendorf tube, 1mL chilled Trizol reagent was added and the solution was drawn through a 21 gauge needle four times and stored at 4°C for 5 minutes. Chloroform (200μL) was added, followed immediately with 30 seconds of vigorous manual agitation, the solution was then stored at 4°C for 5 minutes and then centrifuged at 12 000g for 15 minutes at 4°C. The supernatant was transferred to a new DEPC-treated tube, an equal volume of room temperature isopropanol was added and the tube was stored at -20°C for 30 minutes, centrifuged at 12 000g for 10 minutes at 4°C and the supernatant discarded. The resultant pellet was washed with 70%v/v ethanol (in DEPC diH_2O) and centrifuged at 7 500g for 5 minutes at 4°C (twice). The washed pellet was air dried on ice for 10 minutes, resuspended in 40 to 80μL (decision based on precipitated pellet size) DEPC diH_2O with vigorous vortexing and stored at -80°C until required (<3 months).

2.3.5 Quantitation of Nucleic Acids

Concentration of nucleic acids in solution was determined by UV spectrophotometry using a Beckman DU 640 spectrophotometer. Nucleic acid solutions were diluted to an appropriate concentration with diH_2O and absorbance at 260nm (A_{260}) was measured in duplicate for each sample.

The following formulae were used to calculate concentration;

[DNA]: 1x A_{260} unit = 50ng.μL^{-1}
[RNA]: 1x A_{260} unit = 40ng.μL^{-1}

2.3.6 Quality of Nucleic Acids

The quality of extracted nucleic acids (RNA and DNA) was determined by both UV spectrophotometry and agarose gel electrophoresis.

The ratio of absorbance at 260nm over absorbance at 280nm ($A_{260/280}$) was measured (as described above) for each sample. This gives an indication of contamination with co-extractives (phenol, proteins) with an $A_{260/280}$ ratio greater than 1.8 indicating that the sample is relatively free from contamination. Only nucleic acid solutions with an $A_{260/280}$ ratio greater than 1.8 were selected for use.

Agarose gel electrophoresis gives an indication of the quality and integrity of extracted nucleic acids. Extracted genomic DNA displays a large high molecular weight single band (Figure 2.2: lanes 1-4). Extracted total RNA shows distinct bands corresponding to 28s and 18s rRNA subunits with a multitude of faint bands sized between 1 and 4kb corresponding to mRNA (Figure 2.2: lanes 5-8). Only nucleic acid solutions displaying optimal integrity were selected for use.

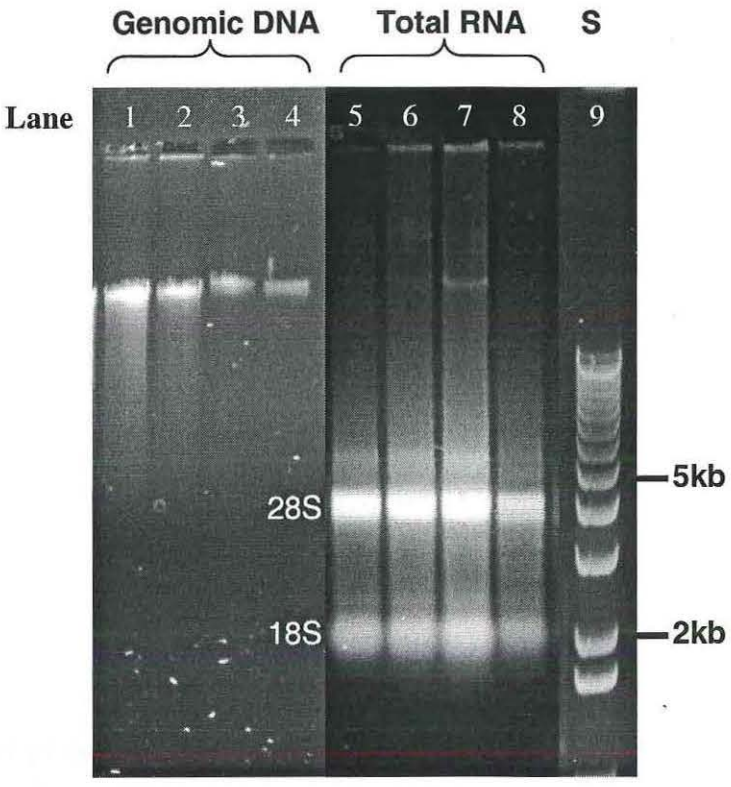


Figure 2.2: Ethidium bromide stained agarose gel showing genomic DNA (lanes 1-4) extracted from *C. destructor* claw closer muscle and total RNA (lanes 5-8) extracted from *C. destructor* embryos. Indicated are the 28S and 18S rRNA subunits. Also shown is size standard ladder (S; lane 9) with sizes indicated in kilobases (kb).

2.3.7 Agarose Gel Electrophoresis of Nucleic Acids

Agarose gel electrophoresis was performed to assess the quality of extracted DNA and RNA, to assess the products of PCR and RT-PCR reactions, and to purify PCR/RT-PCR amplified nucleic acid solutions prior to sequencing.

Agarose gel electrophoresis was performed with either 1%w/v or 3%w/v DNA grade agarose (Progen) in TAE buffer (40mM Tris-acetate pH 8, 1mM EDTA) with 0.5μg.mL^{-1} Ethidium bromide (Sigma). Nucleic acid samples were diluted in 6x loading buffer (50mM EDTA, 40%v/v glycerol, 0.24%w/v Bromophenol blue) and run at constant 110V. Gels were viewed on a UV transilluminator and photographed using a Kodak digital camera.

Size of extracted nucleic acids and amplified products were estimated by comparison to a nucleic acid size standard ladder (1 Kb Plus DNA Ladder; Invitrogen).

2.4 Genetic Analysis

2.4.1 Primer Design for PCR/RT-PCR Amplification of *pgIII* Sequences

Nested degenerate primers were designed to amplify *pgIII* sequences from *C. destructor* DNA and/or RNA. Six *pgIII* genes/cDNAs have been sequenced from other arthropods; *D. melanogaster* (fruit fly; Hexapoda) *Prd*, *Gsb*, and *Gsb-n* (Kilcherr, 1986; Baumgartner *et al.*, 1987); *Schistocerca americana* (grasshopper; Hexapoda) *Pby1* and *Pby2* (Davis *et al.*, 2001); and *Tetranychus urticae* (spider mite; Chelicerata) *Tupax3/7* (Dearden *et al.*, 2002). These sequences, as well as the nearest common ancestor of arthropod and vertebrate *pgIII* genes, *PaxD* from the coral *Acropora millepora* (Cnidaria; Anthozoa) (Miller *et al.*, 2000), were used to predict suitable primer-binding sites for PCR/RT-PCR amplification of *C. destructor pgIII* gene(s)/cDNA(s) and to model evolutionary nucleotide change.

Multiple alignments of the above seven *pgIII* genes/cDNAs were generated in the Multalin program (Corpet, 1988) using the DNA 5-0 selection algorithm. From these alignments areas of high nucleotide homology in the 5′ region encoding the PD and the 3′ region encoding the HD were visually selected as sites for potential primer location. From these homologous sites, nineteen potential degenerate primers were designed. Potential primer sequences were analysed for thermodynamic reliability (self-complementarity, GC content and mean and minimum melting temperatures) using Oligonucleotide properties calculator version 3.02 (Northwestern University, 2002). Primer sequences were also assessed for potential mismatch binding to sequences in GenBank using both FASTA (Pearson, 1990) and BLASTn (Atschul *et al.*, 1997). From these analyses five degenerate primers were selected, two from the 5′ region encoding the PD and three from the 3′ region encoding the HD (Figure 2.3).

2.4.2 Primers

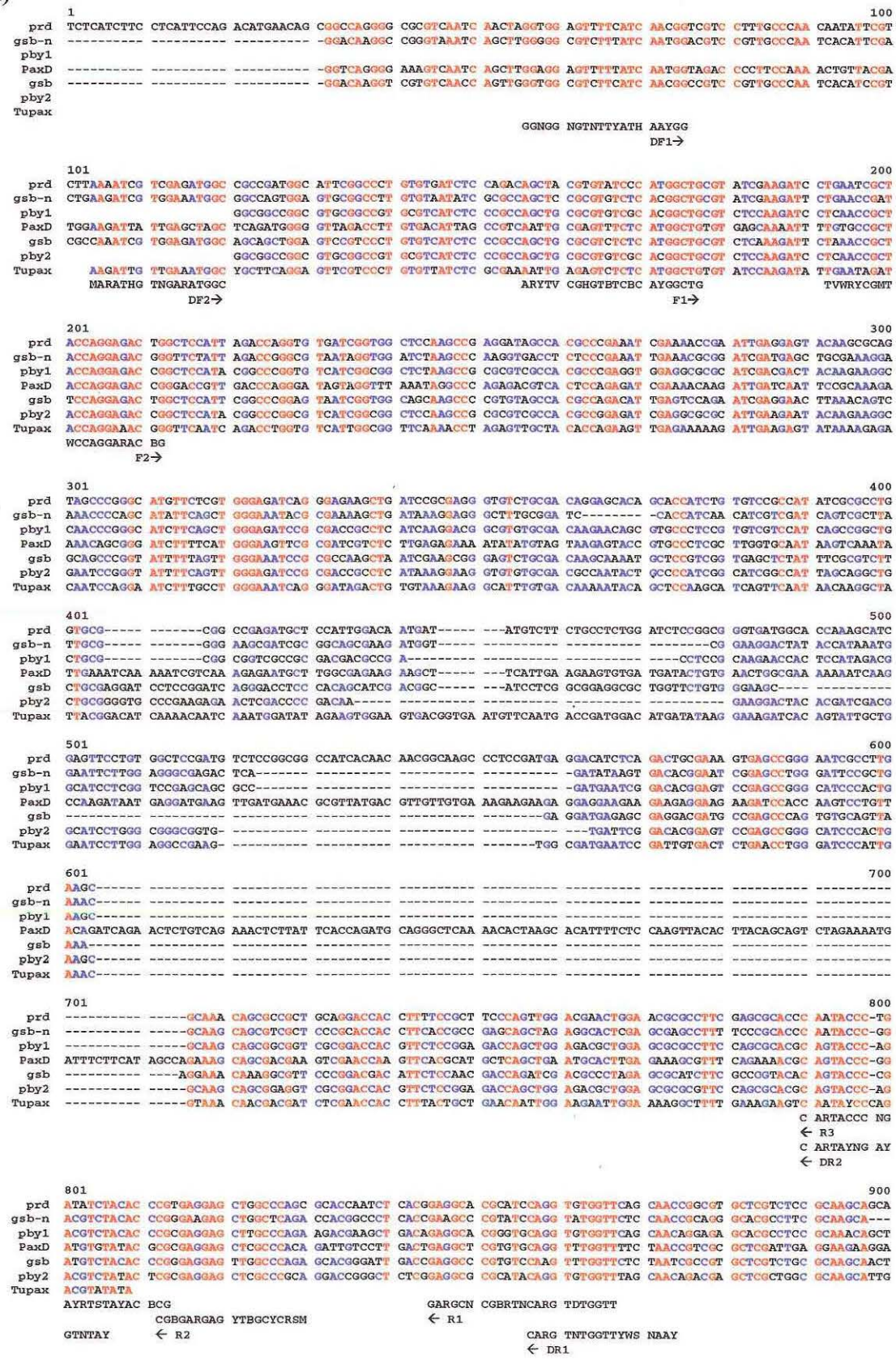
The primers designed and used in this study are shown in Table 2.1. Primers are designated by their direction and their location, thus F1 is a forward primer that is located 5' of F2, and R1 is a reverse primer that is located 5' of R2 and R3. Nested degenerate primers designed by Davis, Jaramillo and Patel (2001) for the isolation of *pgIII* genes from the grasshopper *S. americana* were also used in this study. These primers are given the same designation as above except that they are prefixed with a D, designating Davis *et al.* (2001).

Primers were purchased from Geneworks (Adelaide) and upon arrival in lyophilised form they were immediately diluted to a 100ng.μL^{-1} working solution in DEPC diH_2O, and stored in 50μL aliquots at -20°C.

Table 2.1: Primers used for the PCR/RT-PCR amplification of *pgIII* sequences from *C. destructor*. Primers DF1/2 and DR1/2 were designed by Davis *et al.* (2001). For degenerate code notation see *List of Abbreviations*.

Name	Sequence
F1	5'- ARY TVC GHG TBT CBC AYG GCT G - 3'
F2	5' - TVW RYC GMT WCC AGG ARA CBG - 3'
R1	5' - AAC CAH ACY TGN AYV CGN GCY TC - 3'
R2	5' - KSY GRG CVA RCT CYT CVC G - 3'
R3	5' - CGV GTR TAS AYR TCN GGG TAY TG - 3'
DF1	5' - GGN GGN GTN TTY ATH AAY GG - 3'
DF2	5' - MAR ATH GTN GAR ATG GC - 3'
DR1	5' - RTT NSW RAA CCA NAC YTG - 3'
DR2	5' - RTA NAC RTC NGG RTA YTG - 3'

(A)


prd
gsb-n
pby1
PaxD
gsb
pby2
Tupax
GGNGG NGTNTTYATH AAYGG
DF1→
F1→
DF2→
ARYTV CGHGTBTCBC AYGGCTG
TVWRYCGMT
MARATHG TNGARATGGC
WCCAGGARAC BG
F2→
GARGCN CGBRTNCARG TDTGGTT
← R1
CGBGARGAG YTBGCYCRSM
GTNTAY
← R2
CARG TNTGGTTYWS NAAY
← DR1
C ARTACCC NG
← R3
C ARTAYNG AY
← DR2

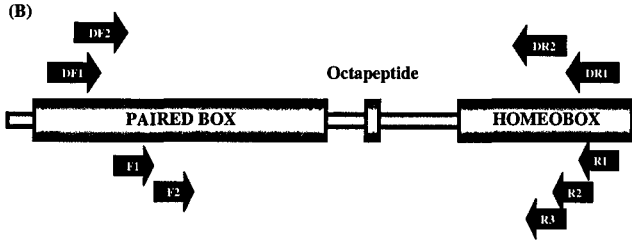


Figure 2.3: Primer binding site locations. (A) Alignment of arthropod *Pax* group III gene/cDNA sequences (*D. melanogaster Prd*, *Gsb*, and *Gsb-n*, *S. americana Pby1* and *Pby2*, *T. urticae Tupax3/7*, and *A. millepora PaxD*; see Table 2.2 for accession numbers) showing binding sites of primers used for the isolation of pgIII sequences from *C. destructor*. Red nucleotides indicate complete identity, blue nucleotides indicate >60% identity, black nucleotides indicate <60% identity. Reverse primers are given as their reverse complement (B) Schematic representation of primers on a representative *Pax* group III gene/mRNA, showing the conserved DNA binding paired box (PD) and homeobox (HD).

2.4.3 Controls

Positive Controls

DNA isolated from whole adult *D. melanogaster* and RNA isolated from a mixture of stage 4 and stage 14 *D. melanogaster* embryos were used as positive controls. Observed amplification products were compared to expected sizes of the amplification products of *Prd*, *Gsb* and *Gsb-n* (from DNA and RNA) to assess specificity of the primers and success of the technique for amplification of *pgIII* genes.

Negative Controls

Control reactions without nucleic acid template were included as a blank in every experiment to assess nucleic acid contamination. Experimental controls containing every ingredient minus either the forward or reverse primer were also conducted for each primer pair, to ensure that the observed amplification products were in fact the product of the primer pair.

Expression Controls

The possibility of co-extraction of DNA with RNA cannot be eliminated, and this was assessed by PCR of all RNA samples (i.e. - in the absence of reverse transcriptase enzyme RNA would not be amplified, thus any observed amplification products would be products of contaminating DNA). Expression controls were conducted for each RT-PCR experiment.

2.4.4 Polymerase Chain Reaction (PCR)

PCR was performed using HotStarTaq DNA polymerase (Qiagen). PCR reactions were performed using an MJ Research PTC-200 Peltier thermal cycler with a heated lid. All samples were set up on ice.

Optimal conditions for PCR were; reaction volumes of 20μL containing final concentrations of $4\mu g.mL^{-1}$ DNA template (80ng DNA per reaction), 1μM each primer, 400μM each dNTP (Finnzymes), 15mM $MgCl_2$, 1x PCR buffer (Qiagen), 1x Q solution (Qiagen), 0.5 units HotStarTaq DNA polymerase (Qiagen). Optimal cycling conditions were; an enzyme activation step of 95°C for 15 minutes, followed by 45 cycles of; 94°C for 1 minute (denaturation), 56°C (primers F1/2, R1/2/3) or 50°C (DF1/2, DR1/2) for 1 minute (annealing), 72°C for 1 minute (extension); then a final extension at 72°C for 10 minutes.

2.4.5 Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR)

RT-PCR was performed using the OneStep RT-PCR Kit (Qiagen) containing recombinant heterodimeric Omniscript and Sensiscript reverse transcriptase enzymes and HotStartTaq DNA polymerase. RT-PCR reactions were performed using an MJ Research PTC-200 Peltier thermal cycler with a heated lid. Samples were set up on ice and loaded into a pre-heated machine (50°C).

Optimal conditions for RT-PCR were; reaction volumes of 25μL containing final concentrations of $5\mu g.mL^{-1}$ total RNA template (125ng total RNA per reaction), 1μM each primer, 400μM each dNTP (Qiagen), 12.5mM $MgCl_2$, 1x RT-PCR buffer (Qiagen), 1x Q

solution (Qiagen), 1.0μL OneStep RT-PCR enzyme mix (Qiagen). Optimal cycling conditions were; a reverse transcription reaction at 50°C for 30 minutes, a HotStarTaq polymerase enzyme activation step of 95°C for 15 minutes, followed by 45 cycles of; 94°C for 1 minute (denaturation), 56°C (primers F1/2, R1/2/3) or 50°C (DF1/2, DR1/2) for 1 minute (annealing), 72°C for 1 minute (extension); then a final extension at 72°C for 10 minutes.

2.4.6 Nested Reactions

As degenerate primers were used for this analysis, it proved necessary to use a nested technique to amplify enough specific product from the RT-PCR reactions. Nested reactions utilise two separate reactions each with a distinct set of primers. This increases both the specificity of a particular amplification and the amount of specific product obtained. Nested PCR/RT-PCR involves two rounds of reactions. The first round involves PCR/RT-PCR amplification of template DNA/RNA. An aliquot of the first round reaction is then used as the template for the second round. The second round reaction uses forward and reverse primers that sit internally of the forward and reverse primers used in the first reaction (Figure 2.4; A), and so only amplifies products generated in the first round that also contain binding sites for the second round primers. The second round amplification is highly specific to the gene(s) of interest as the only template for the second round is that already amplified by the primers used in the first round (Figure 2.4).

In the optimal RT-PCR nested reaction, primers F1 and R1 were used in the first round (RT-PCR) to produce cDNAs from *C. destructor* embryos/muscle samples or control RNA (*D. melanogaster* embryo RNA). The first round cDNAs were then amplified by primers F2 and R3 in the second round (nested PCR). Optimal conditions for nested PCR reactions were; reaction volumes of 20μL containing final concentrations of 5%v/v first round RT-PCR reaction solution (containing cDNA template), 1μM each primer, 400μM each dNTP (Finnzymes), 15mM $MgCl_2$, 1x PCR buffer (Qiagen), 1x Q solution (Qiagen), 0.5 units HotStarTaq DNA polymerase (Qiagen). Optimal cycling conditions were; an enzyme activation step of 95°C for 15 minutes, followed by 35 cycles of; 94°C for 1 minute (denaturation), 56°C (using primers F2 and R3) for 1 minute (annealing), 72°C for 1 minute (extension); then a final extension at 72°C for 10 minutes.

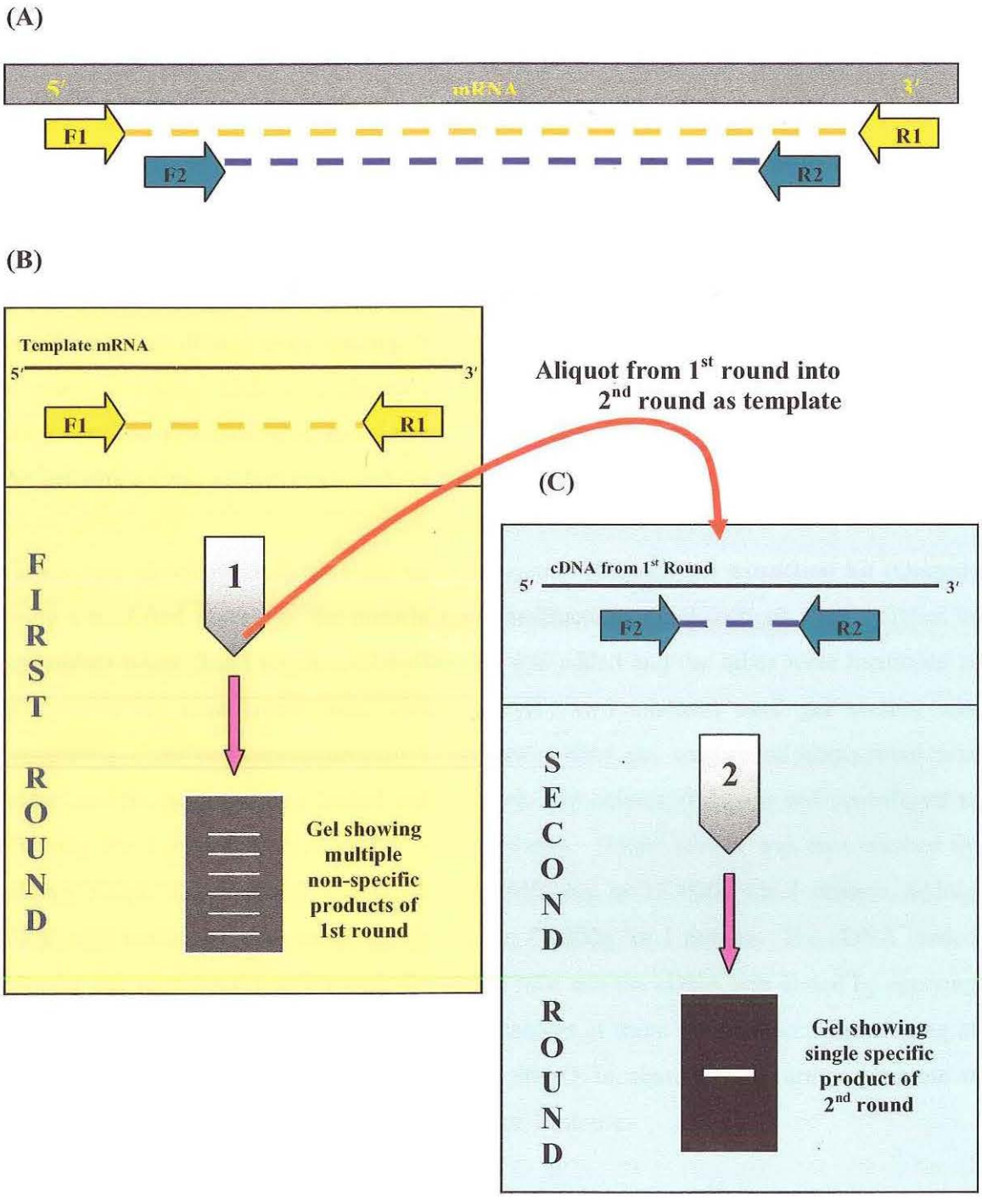


Figure 2.4: Schematic representation of nested reaction. Shown are (A) respective locations of first round primers (F1 and R1) and internally located second round primers (F2 and R2) on target mRNA, and (B) first round reaction showing target mRNA with primers and resultant gel showing multiple non-specific product bands, and (C) second round reaction showing target cDNA (aliquot from first round reaction) with internal primers and resultant gel showing single specific product band.

2.4.7 Gel Purification of RT-PCR Reactions

Nested RT-PCR reactions performed on *C. destructor* embryo RNA consistently produced two bands (two cDNAs). Thus direct sequencing was not possible, as the sequence would be an aggregate of both sequences. To overcome this, nested RT-PCR reaction solutions were gel purified, so that each band (cDNA) could be extracted and sequenced directly. Nested RT-PCR products from *C. destructor* embryos or regenerating muscle samples (100 or 200μL) were diluted in 6x loading buffer and run on a 200mm long 1%w/v agarose gel at constant 110V. Gels were placed on a UV transilluminator covered in two layers of cling film and illuminated on low power. Each nested RT-PCR product was excised from the gel with a fresh scalpel blade and loaded into an individual 1.7mL Eppendorf tube.

cDNA was eluted from agarose gel sections by the Qiaquick gel extraction kit (Qiagen) using a modified version of the manufacturers instructions. Gel sections were weighed in Eppendorf tubes, 3 gel volumes of buffer QG was added and the tubes were incubated at 50°C in a dry heat block (with vortexing every two minutes) until gel section was completely dissolved (approximately 7 minutes). One gel volume of isopropanol was added and this solution was loaded onto a quick-spin column (Qiagen) and centrifuged at 13 000g for 1 minute to bind cDNA to the column. Bound cDNA was then washed by adding 300μL QG buffer to the column, centrifuging at 13 000g for 1 minute, adding 800μL QE buffer and then centrifuging twice at 13 000g for 1 minute. The cDNA loaded column was then fitted into a 1.7mL Eppendorf tube and the cDNA was eluted by applying 30μL diH_2O to the column, incubating for 3 minutes at room temperature, centrifuging at 13 000g for 1 minute, applying a further 20μL diH_2O, incubating for a further 1 minute at room temperature and centrifuging at 13 000g for 1 minute.

Eluted cDNA solution was quantified and checked for contamination by UV spectrophotometry, adjusted to a cDNA concentration of 10ng.μL^{-1} and stored at -20°C until required (<2 weeks).

2.4.8 Sequencing of Gel Purified cDNA

Gel purified *C. destructor* cDNA solutions, at a concentration of 10ng.μL^{-1}, and primer solutions, at a concentration of 1pmol.μL^{-1}, were frozen and sent off to the Western Australian Genome Resource Centre (Department of Clinical Immunology and Biochemical Genetics, Royal Perth Hospital) and sequenced on an Applied Biosystems ABI 3100 16 capillary automated DNA sequencer using ABI BigDye Terminator version 3.1 chemistry.

Sequences were analysed from electropherograms using Chromas version 2.23 (2002). Aggregate consensus sequences were determined from alignments using Multalin (Corpet, 1988) of four forward and five reverse reactions for each cDNA product from *C. destructor* embryos and from *C. destructor* regenerating muscle and limb.

2.4.9 Sequence Analysis of *Pax* Group III Sequences

Sequences were analysed for homology with all known genes in GenBank using both BLASTn and BLASTx (nucleotide and translated basic local alignment search tool, respectively; Altschul *et al.*, 1997), from which open reading frames and translations were deduced. Sequences were aligned with known pgIII protein sequences using Multalin (Corpet, 1988) and ClustalW (Higgins *et al.*, 1994).

2.4.10 Phylogenetic Analysis

Phylogenetic analysis was performed upon the deduced partial protein sequences obtained for the PD (86 amino acids) and HD (48 amino acids) of an isolated *C. destructor pgIII* cDNA. This method reduces the potential confounding influences of both homoplasy (convergent alterations and reversals of amino acids) and neutral mutation (nucleotide mutations that do not alter the amino acid sequence) and thus is purported to be a more accurate reflection of functionally important relationships between orthologous and paralogous genes (Agosti *et al.*, 1996; Felsenstein & Sober, 1986). Performing analyses on the protein sequence eliminates any neutral mutations, and the effects of homoplasy are reduced by analysing the most slowly evolving regions of the protein (Freeman & Herron,

2004). The DNA-binding regions of Pax proteins are highly conserved, mutations in this region may be of functional significance (Apuzzo & Gros, 2002; Sun *et al.*, 2002; Xue *et al.*, 2001; Kozmik *et al*, 1997; Bertuccioli *et al.*, 1996), whereas the region separating these DNA-binding domains is highly variable (Breitling & Gerber, 2000; Sun *et al.*, 1997).

Maximum likelihood analysis was conducted upon 86 amino acids of the PD and 48 amino acids of the HD of *C. destructor* pgIII deduced protein sequence, and other representative Pax proteins (given in Table 2.2). Sequences were aligned using ClustalW (Higgins *et al.*, 1994) and maximum likelihood analysis was performed in the ProML (protein maximum likelihood) program of PHYLIP (2002) using the Jones-Taylor-Thornton (1992) model of amino acid change for branch length estimation. Phylogenetic trees (phylograms) were constructed from the results of maximum likelihood analysis. Branch lengths in phylograms are proportional to amino acid change, thus the longer the branch between two proteins, the more divergent their amino acid sequence.

Statistical analysis of the branching patterns given by maximum likelihood analysis was performed by bootstrapping. Bootstrapping was performed on protein sequence data using the Seqboot program in PHYLIP, generating 100 pseudoreplicate analyses. Bootstrap values represent the percentage of pseudoreplicates in which the given branch was reconstructed, and therefore bootstrap values give a probability rating of the positioning of the branch of the phylogram that they specify. The greater the bootstrap value, the greater the certainty of branch positioning, with bootstrap values less than 40% indicating uncertainty in the position of the specified branch (Miller *et al.*, 2000). The extended majority rule consensus phylogram was determined using the Consense program in PHYLIP. Maximum likelihood phylograms were constructed in TreeView (Page, 1996).

Table 2.2: Representative Pax proteins used for maximum likelihood analysis.

Pax group	Protein	Species	Accession number (GenBank)
I	Pax1	*Mus musculus*	M69222
	Pax9	*M. musculus*	P47242
	Amphipax1	*Branchiostoma floridae*	U20167
	Pox-meso	*Drosophila melanogaster*	X16992
II	Pax2	*M. musculus*	NM011037
	Pax5	*M. musculus*	Q02650
	Pax8	*M. musculus*	Q00288
	Amphipax2	*B. floridae*	AF053762
	HrPax-258	*Halocynthia roretzi*	AB006675
	PaxB	*Acropora millepora*	AF241310
	Sparkling	*D. melanogaster*	AF010256
III	Pax3	*M. musculus*	X59358
	Pax7	*M. musculus*	P47239
	Amphipax3/7	*B. floridae*	AF165886
	PaxD	*A. millepora*	AF241311
	Tupax3/7	*Tetranychus urticae*	AY148194
	Pairberry 1 (Pby1)	*Schistocerca americana*	AY040535
	Pairberry 2 (Pby2)	*S. americana*	AY040536
	Paired (Prd)	*D. melanogaster*	M14548
	Gooseberry (Gsb)	*D. melanogaster*	NM079139
	Gooseberry-neuro (Gsb-n)	*D. melanogaster*	NM079138
IV	Pax4	*M. musculus*	P32115
	Pax6	*M. musculus*	X63963
	Amphipax6	*B. floridae*	AJ223444
	PaxA	*A. millepora*	AF053458
	PaxC	*A. millepora*	AF053459
	Pox-neuro	*D. melanogaster*	X58917
	Eyeless	*D. melanogaster*	X79493

3.0 RESULTS

3.1 Introduction

To isolate *pgIII* gene sequences from *C. destructor*, PCR was conducted on genomic DNA and RT-PCR was conducted on RNA isolated from embryos. PCR/RT-PCR primers were designed to complement conserved regions of known *pgIII* sequences from evolutionary relatives of *C. destructor*. The ability of these primers to amplify *pgIII* sequences by PCR and RT-PCR was first validated on a positive control (*D. melanogaster* DNA and embryo RNA). Parameters of the PCR and RT-PCR methodology were optimised for the amplification of *pgIII* sequences from *D. melanogaster* DNA and embryo RNA. Once validated and optimised, this technique was then applied to the isolation of *pgIII* sequences from *C. destructor*. PCR was conducted on *C. destructor* genomic DNA and RT-PCR was conducted on *C. destructor* embryos at 30% and 50% of development.

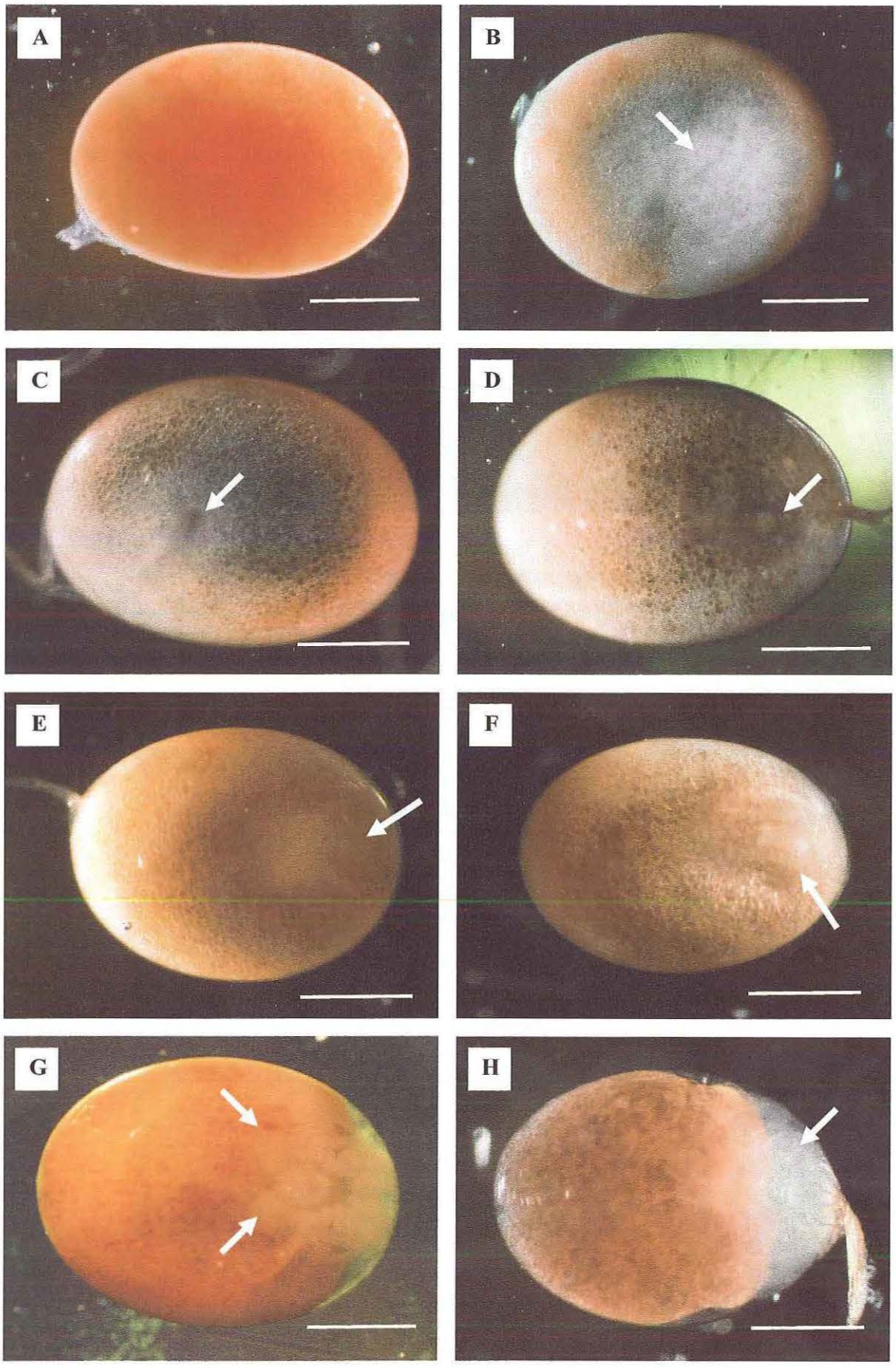
The products of RT-PCR from *C. destructor* embryos were gel-purified and sequenced. The sequences obtained from *C. destructor* were analysed for homology to known genes through searches of GenBank. The deduced protein sequence of an isolated *C. destructor pgIII* cDNA was then compared to the sequences of other Pax proteins.

As vertebrate *Pax7* specifies adult myoblasts (satellite cells) and is required for adult muscle regeneration (Seale *et al.*, 2000), and *C. destructor* has a phenomenal capability to regenerate muscle as an adult (Cutler *et al.*, 2002), *pgIII* gene expression was also assessed in regenerating muscle and limb of *C. destructor* by RT-PCR.

3.2 Embryonic Development in *C. destructor*

The external morphology of embryonic development in *C. destructor* has been previously studied by Sandeman and Sandeman (1991). In this research it was necessary to confirm their results and ensure that embryos used in this study were at the desired stages (30% and 50% of development). Embryonic development (extrusion to hatching) took 39 days at a water temperature of 19°C. Freshly extruded embryos were approximately 2.7mm in length with a breadth of 1.7mm, and did not change appreciably in size throughout embryonic

development. The chorion of *C. destructor* embryos is thin and transparent, and so was not removed for observation. Newly laid embryos were filled with a heterogeneous yolky matrix containing cellular material (blastoderm cells; Figure 3.1: A). At 10% (4 days; Figure 3.1: B) an aggregation of cells can be seen on the ventral surface (the incipient germinal disc), by 20% (8 days; Figure 3.1: C) these ventral cells have sunk into the egg, forming an oval cluster of cells on the egg surface. At 30% of development (12 days; Figure 3.1: D) a dark area surrounds the embryo (a white cluster of cells), this signifies that the embryo is in the nauplius stage (Sandeman & Sandeman, 1991). At 40% of development (16 days; Figure 3.1:E) the dark area surrounding the embryo now covers more than half of the ventral side of the egg, and post-naupliar structures can be discerned on the white embryo. Embryos at 50% of development (20 days; Figure 3.1: F) are distinguished by the first appearance of cheliped and other limb buds (Figure 3.2). Embryos at 60% of development (24 days; Figure 3.1: G) are distinguished by the differentiation of the eyes into lobes. Embryos at 80% of development (32 days; Figure 3.1: H) have fully elongated chelipeds and other appendages, and, following hatching, emerge as miniature adults in post-embryonic stage 1 (POI) (40 days; Figure 3.1: I).


A
B
C
D
E
F
G
H

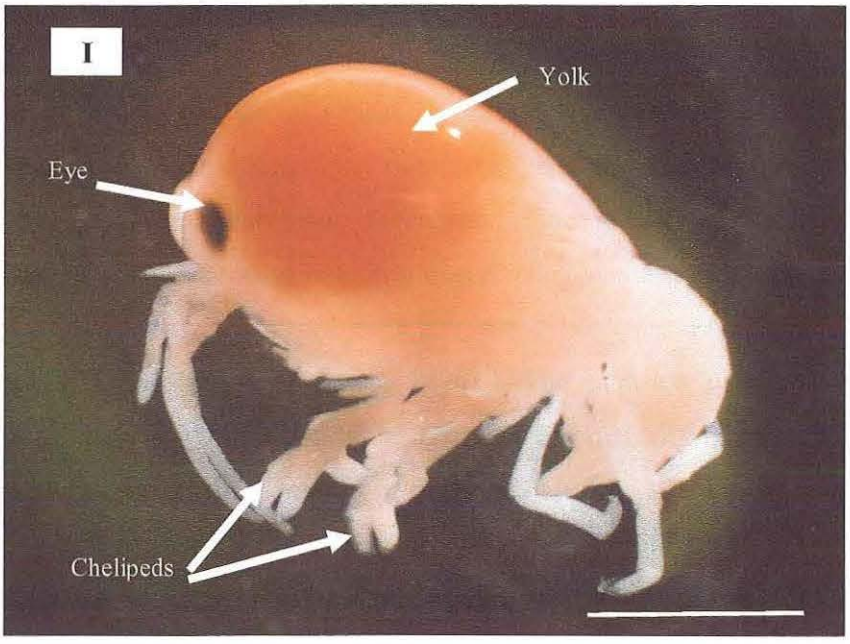


Figure 3.1: Stages in the embryonic development of *C. destructor*. (A) freshly extruded embryos (<10%); (B) 10% (arrow indicates incipient germinal disc); (C) 20% (arrow indicates oval cluster); (D) 30% (arrow indicates nauplius embryo); (E) 40% (arrow indicates embryo); (F) 50% (arrow indicates emerging cheliped; shown at greater magnification in Figure 3.2); (G) 60% (arrows indicate eyes); (H) 80% (arrow indicates embryo appendages); (I) PO1 (1^{st} postembryonic stage; arrows indicate yolk inside carapace, eyes, and chelipeds). *Scale bar*: 1.0mm.

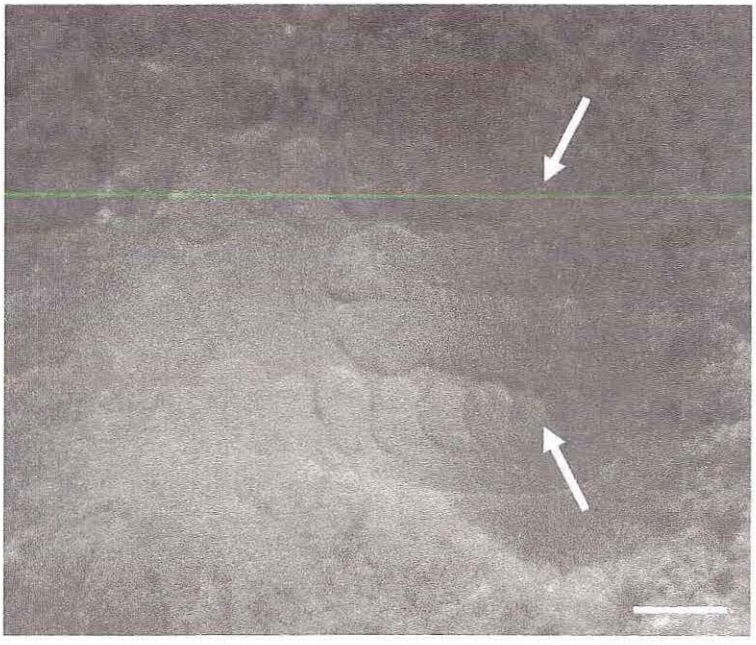


Figure 3.2: *C. destructor* embryo at 50% of development. Arrows indicate emerging cheliped buds (high magnification of Figure 3.1: F). *Scale bar*: 100µm.

3.3 Limb and Muscle Regeneration in *C. destructor*

C. destructor undergoes two distinct types of muscle regeneration. The first type, moult-cycle regeneration, involves the atrophy of muscle prior to moulting (premoult), followed by a moult (ecdysis), and then rapid regeneration and growth of the muscle following ecdysis (postmoult). This process is highly exacerbated in the claw closer muscle, and the muscle regeneration process appears to begin in the late premoult phase (Lamey *et al.*, 2002; West, 1997; Holland & Skinner, 1976). The duration of the moult cycles of experimental animals was highly variable, however the majority of experimental animals moulted three times during the 5 months of experimentation.

The other type of muscle regeneration performed by *C. destructor* is epimorphic regeneration (the regeneration of entire limbs). Epimorphic cheliped regeneration has been studied previously by Cutler *et al.* (2002); in this research it was necessary to confirm their results. Epimorphic regeneration in *C. destructor* occurred in defined stages (Figure 3.3). Several days following cheliped autotomy at the basi-ischial joint breakage plane (Figure 3.3; A and B), the wound sealed off with a blastema (Figure 3.3: C). The blastema then erupted, eventually extending to form an unsegmented papilla (Figure 3.3: D). This papilla then extended further, became segmented and differentiated (the closed dactyl stage; Figure 3.3: E).

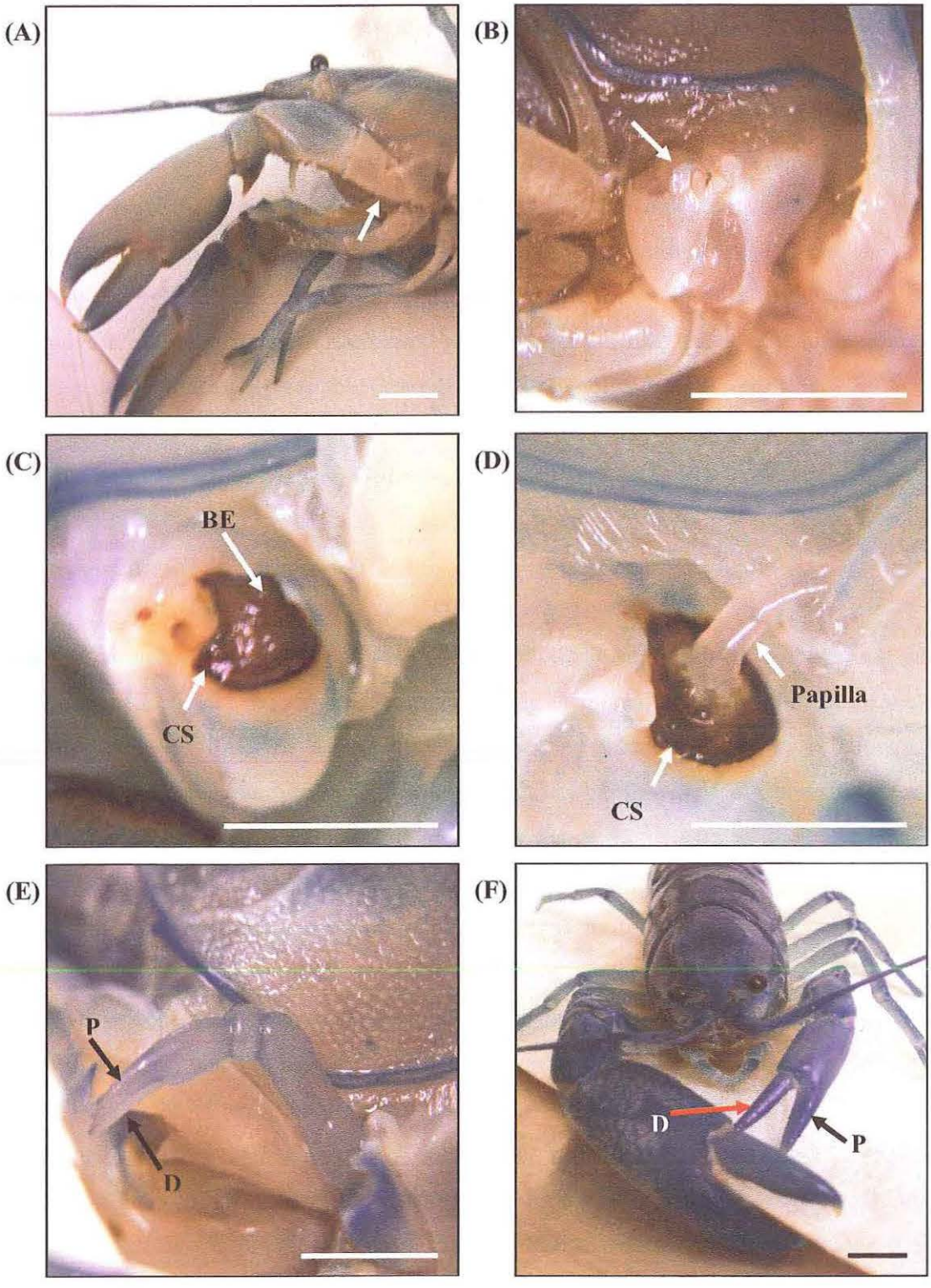


Figure 3.3: Epimorphic cheliped regeneration in *C. destructor*. (A) pre-autotomy cheliped (arrow indicates basi-ischial breakage plane), (B) post-autotomy stump (indicated by arrow), (C) blastema eruption (CS: coxal stump; BE: blastema eruption), (D) unsegmented papilla (CS: coxal stump; and papilla), (E) the segmented regenerate at the closed dactyl stage (P: propus; D: dactyl, which are fused), and (F) the fully regenerated functional limb post moult (P: propus; D: dactyl, which are separated). *Scale bar*: 5.0mm.

At the late closed dactyl stage, muscle regeneration is apparently completed, and fibres have been demonstrated to be contract *in vitro* (West *et al.*, 1995). After the closed dactyl stage growth and differentiation was completed, the externally observable limb regeneration process appeared to cease (while the animal prepares to moult). Following ecdysis a small fully-functional replica of the original claw emerged (Figure 3.3: F). At a water temperature of 19°C, the period of epimorphic regeneration was between 15 and 20 days to the undifferentiated papilla stage, and approximately 30 days to the closed dactyl stage. The duration between the closed dactyl stage and fully functional limb emergence (after ecdysis) was highly variable (linked to the timing of ecdysis).

3.4 PCR and RT-PCR Method Optimisation

3.4.1 Method Optimisation

As degenerate primers were used for PCR and RT-PCR, and each different primer within the degenerate set has a different annealing temperature, the specific annealing temperatures could not be accurately determined by thermodynamic calculations. For example, the degenerate primer CCR would contain both primers CCA and CCG, each of which would have different specific annealing temperatures due to their different GC content. As we do not know in advance which of these primers will be gene-specific, then calculating their respective annealing temperatures is futile. Thus optimal stringency and conditions were determined empirically. Optimal nucleic acid template (DNA, RNA, cDNA) concentration, primer concentration, annealing temperatures, number of cycles, extension times, and the use of Q solution (Qiagen) were all determined for each primer pair (Table 3.1). The optimal conditions were first determined for positive controls (*D. melanogaster* DNA and embryo RNA) and then applied and adjusted for *C. destructor* DNA and embryo RNA samples. Selection criteria for determining optimal conditions for positive controls were the lowest stringency conditions without production of non-specific bands (amplification product other than *pgIII* genes). The optimal conditions are given in Table 3.1.

Table 3.1: Parameters varied during PCR and RT-PCR method optimisation.

Parameter	Variation	Optimal
Template concentration	DNA: 0.4, 1, 4, 10, & 40μg.mL^{-1}	4μg.mL^{-1} DNA
	RNA: 0.5, 1, 5, 10, & 50μg.mL^{-1}	5μg.mL^{-1} RNA
	cDNA (1st round): 0.5, 5, & 10%v/v	5%v/v
Primer concentration	0.2, 0.4, 0.6, 0.8, 1.0, & 2.0μM	1μM each primer
Q solution	Presence/absence	Present
Annealing temperature	50, 54, 56, 58, 60, & 65°C	50°C (DF1/2, DR1/2)
		56°C (F1/2, R1/2/3)
Number of cycles	30, 35, 40, 45, & 50 cycles	45 (1st round)
		35 (2nd round)
Extension time	45s, 1, 2, & 3 minutes	1 minute

All nested reactions were performed under optimised PCR conditions (below), with the exception of nucleic acid template amount and number of cycles. These two conditions were determined empirically. Template concentrations tested were 0.5, 5, or 10%v/v of the final reaction volume. The number of cycles of the nested reaction was varied between 25 to 40 cycles. Primer combinations tested included; 1st round: DF1/DR1, followed be 2nd round: DF2/DR2, F1/R1, F2/R2, or F2/R3, and 1st round: F1/R1, followed by 2nd round: F2/R2, or F2/R3. Optimal conditions are given in Table 3.1.

3.4.2 Optimised Conditions for PCR

The optimised conditions for this reaction were; reaction volumes of 20μL containing final concentrations of 4μg.mL^{-1} *D. melanogaster* DNA template (80ng DNA per reaction), 1μM each primer, 400μM each dNTP, 15mM $MgCl_2$, 1x PCR buffer, 1x Q solution, 0.5 units HotStarTaq DNA polymerase. Optimal cycling conditions were; an enzyme activation step of 95°C for 15 minutes, followed by 45 cycles of; 94°C for 1 minute (denaturation), 56°C for 1 minute (annealing), 72°C for 1 minute (extension), then a final extension at 72°C for 10 minutes.

3.4.3 Optimised Conditions for RT-PCR

The optimal conditions for nested RT-PCR were as follows. First round (RT-PCR); reaction volumes of 25μL containing final concentrations of 5μg.mL^{-1} total RNA template (125ng total RNA per reaction), 1μM each primer (F1 and R1), 400μM each dNTP, 12.5mM $MgCl_2$, 1x RT-PCR buffer, 1x Q solution, 1.0μL OneStep RT-PCR enzyme mix. Optimal cycling conditions were; a reverse transcription reaction at 50°C for 30 minutes, a HotStarTaq polymerase enzyme activation step of 95°C for 15 minutes, followed by 45 cycles of; 94°C for 1 minute (denaturation), 56°C for 1 minute (annealing), 72°C for 1 minute (extension), then a final extension at 72°C for 10 minutes. Second round (nested PCR); reaction volumes of 20μL containing final concentrations of 5%v/v first round RT-PCR reaction solution (1μL aliquot), 1μM each primer (F2 and R3), 400μM each dNTP, 15mM $MgCl_2$, 1x PCR buffer, 1x Q solution, 0.5 units HotStarTaq DNA polymerase. Optimal cycling conditions were; an enzyme activation step of 95°C for 15 minutes, followed by 35 cycles of; 94°C for 1 minute (denaturation), 56°C for 1 minute (annealing), 72°C for 1 minute (extension); then a final extension at 72°C for 10 minutes.

3.4.4 Expected Sizes of *D. melanogaster pgIII* PCR/RT-PCR Products

Primers designed from alignments of known arthropod *pgIII* gene sequences were first trialled against a positive control; DNA and RNA isolated from *D. melanogaster* whole flies and embryos, respectively. To assess the success of amplification of *pgIII* genes by this process, the size of expected amplification products was first calculated from the alignment of primers against DNA and cDNA sequences of the three *D. melanogaster pgIII* genes; *Prd*, *Gsb*, and *Gsb-n* (Figure 3.4).

Figure 3.4 shows an alignment of primers F1, F2, R1, R2, and R3 against *Prd, Gsb, and Gsb-n* cDNA sequences. *Prd* and *Gsb* do not have introns within the region displayed in Figure 3.4, so this alignment is representative of both cDNA and DNA for these two genes. The *Gsb-n* gene has four introns, two of which are resident within the PCR amplified region (intron 2: 4140bp and intron 3: 6120bp). Expected sizes of *Prd*, *Gsb*, and *Gsb-n* from PCR amplification of DNA, and RT-PCR amplification of RNA, using combinations of primers F1, F2, R1, R2, and R3 are given in Table 3.2.

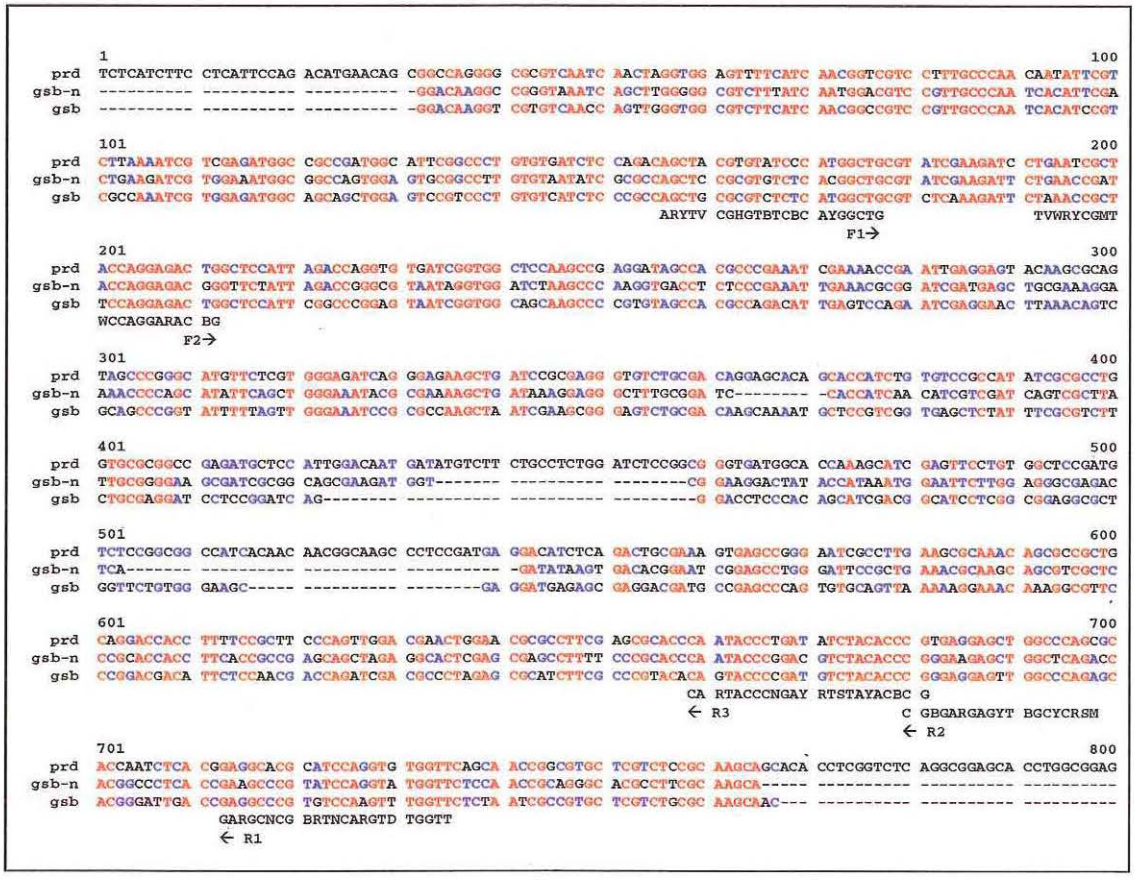


Figure 3.4: Alignment of primers F1, F2, R1, R2, and R3 against the three *D. melanogaster pgIII* cDNA sequences (*Prd*, *Gsb*, and *Gsb-n*) used for calculation of PCR/RTPCR amplification region size. Red nucleotides indicate complete consensus, blue nucleotides indicate >60% consensus, black nucleotides indicate no consensus. Dashes (-) indicate nucleotide deletions. All reverse primers are shown as their reverse complement.

3.4.5 PCR of *D. melanogaster* Genomic DNA

PCR of *D. melanogaster* genomic DNA using combinations of forward and reverse primers generated two products (Figure 3.5). Each of these PCR reactions generated products corresponding to the predicted sizes of *Prd* and *Gsb* (Table 3.2). *Gsb-n* was not amplified by PCR of *D. melanogaster* genomic DNA. The predicted sizes of the amplification of *Gsb-n* from DNA (Table 3.2) all exceed the limits of this technique (approximately 4kb; Qiagen, 2002), and so, no *Gsb-n* product would be expected from PCR of *D. melanogaster* DNA using any of the primers designed for this investigation.

Table 3.2: Predicted (calculated) sizes of *D. melanogaster Prd*, *Gsb*, and *Gsb-n* amplification regions for PCR and RT-PCR and observed product sizes from PCR and nested RT-PCR reactions on *D. melanogaster* genomic DNA and embryo RNA (measured form agarose gel separated reactions shown in Figures 3.5 and 3.6). N/P: no product obtained.

Gene/cDNA	Primers		Genomic DNA length		RNA (cDNA) length	
	Forward	Reverse	predicted	observed	predicted	observed
Prd	F1	R1	580bp	~580bp	580bp	
	F1	R2	543bp	~540bp	543bp	
	F1	R3	526bp	~530bp	526bp	
	F2	R1	544bp	~520bp	544bp	
	F2	R2	507bp	~500bp	507bp	
	F2	R3	490bp	~490bp	490bp	~500bp
Gsb	F1	R1	520bp	~510bp	520bp	
	F1	R2	483bp	~490bp	483bp	
	F1	R3	466bp	~470bp	466bp	
	F2	R1	484bp	~480bp	484bp	
	F2	R2	447bp	~440bp	447bp	
	F2	R3	430bp	~520bp	430bp	~450bp
Gsb-n	F1	R1	10 768bp	N/P	508bp	
	F1	R2	10 731bp	N/P	471bp	
	F1	R3	10 714bp	N/P	454bp	
	F2	R1	10 732bp	N/P	472bp	
	F2	R2	10 695bp	N/P	435bp	
	F2	R3	10 678bp	N/P	418bp	~420bp

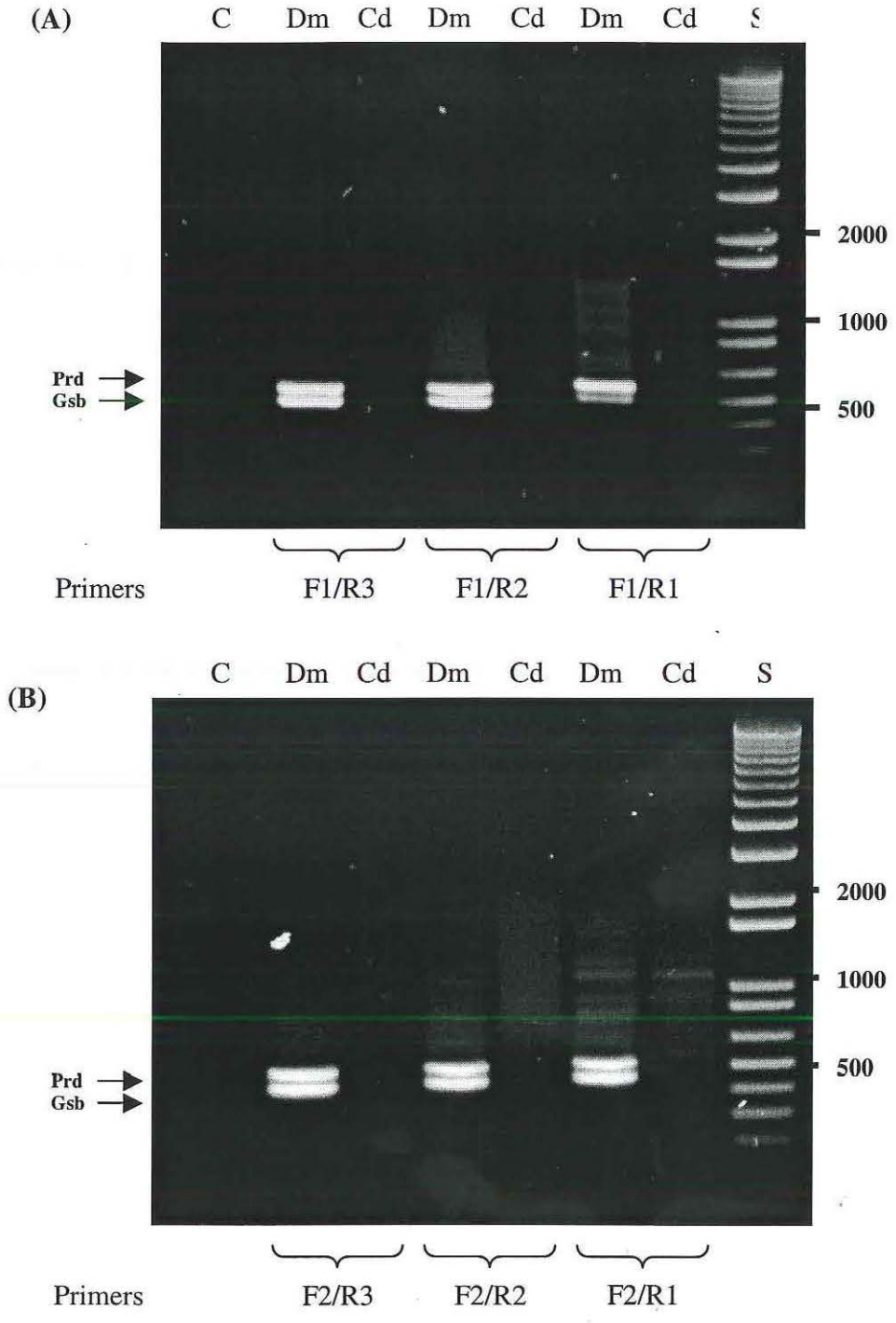


Figure 3.5: Ethidium bromide stained agarose gels showing PCR amplification of *C. destructor* (Cd) and *D. melanogaster* (Dm) genomic DNA using primers F1 with R1/2/3 (A) or F2 with R1/2/3 (B). Bands corresponding to the predicted sizes of *Prd* and *Gsb* (as indicated) are shown. DNA size standard (S) and blank control (C) are also shown.

3.4.6 RT-PCR of *D. melanogaster* Embryo RNA

RT-PCR of RNA isolated from stage 8 and stage 14 *D. melanogaster* embryos using primers F1 and R1 generated a single product of 520bp (Figure 3.6). This product corresponds to the predicted size of *Gsb* (Table 3.2). As *Prd* and *Gsb-n* did not appear to be sufficiently amplified by this RT-PCR reaction, a nested reaction was conducted on the product of this reaction. The nested reaction consisted of a first round RT-PCR amplification using primers F1 and R1, and a second round PCR amplification using primers F2 and R3.

The nested RT-PCR reaction (using primers F1/R1 then F2/R3) amplified three products from *D. melanogaster* embryo RNA (Figure 3.6). These products corresponded to the predicted sizes of *Prd* (490bp), *Gsb* (430bp), and *Gsb-n* (418bp) (Table 3.2).

3.4.7 Other Controls

Three types of controls were used in this methodology. The primers developed in this research were tested by both positive and negative controls. The amplification of *pgIII* genes from *D. melanogaster* genomic DNA and embryo RNA constituted the positive control. As indicated above, these primers correctly amplified *pgIII* sequences from *D. melanogaster*. The negative control consisted of reactions (PCR and RT-PCR of *D. melanogaster* and *C. destructor* genomic DNA and embryo RNA) containing all constituents but only either the forward or reverse primer. None of these reactions generated products, indicating that the products amplified in experimental reactions were products of amplification using both forward and reverse primers.

Negative controls containing all reaction constituents minus nucleic acid template (blank controls) were included in all reactions to control for contamination. These reactions are shown with each agarose gel (Figures 3.5 and 3.6) and do not display amplification products, indicating that the products obtained are not the result of contamination.

In all RT-PCR experiments assaying expression of *pgIII* genes, expression controls were included to ensure that the observed amplification was due to the presence of *pgIII* RNA

(and hence gene expression). These expression controls contain all reaction constituents minus reverse transcriptase enzyme, and hence do not amplify RNA. Expression controls are included for each expression experiment and are run parallel to these reactions in the agarose gel shown in Figure 3.6. None of these expression controls contain product bands, therefore the product bands present in the experimental reactions are due to the presence of RNA, and hence gene expression.

3.5 PCR/RT-PCR Isolation of *C. destructor Pax* Group III Sequences

As the developed primers/techniques had successfully amplified *pgIII* sequences from *D. melanogaster* DNA and embryo RNA, they were then applied to the isolation of *pgIII* sequences from *C. destructor*. Because *pgIII* sequences in *C. destructor* are unknown, the stringency conditions of PCR and RT-PCR (template and primer concentrations; annealing temperatures) were again optimised for DNA and RNA isolated from *C. destructor*.

3.5.1 PCR of *C. destructor* Genomic DNA

PCR using any combination of forward and reverse primers (F1/R1, F1/R2, F1/R3, F2/R1, F2/R2, and F2/R3) did not produce any amplification products from *C. destructor* genomic DNA (Figure 3.5). A small amount of a ~1.1kb product was generated by primers F2/R1 (Figure 3.5), however this product was not observed in any of the primers that sit internally to these two primers (F2/R2 and F2/R3), and thus this product was not a *pgIII* sequence. An increase in extension times to 3 minutes (which extends the possible amplification length of sequences) did not result in any product. Nested PCR was also trialled, using primers F1/R1 in the first round followed by primers F2/R3 in the second round. No amplification products were produced by this reaction.

3.5.2 RT-PCR of *C. destructor* Embryo RNA

Nested RT-PCR using primers F1/R1 (first round) followed by primers F2/R3 (second round) of RNA isolated from *C. destructor* embryos at 30% and 50% of development yielded two amplification products (Figure 3.6). The larger product produced was approximately 420bp and the smaller product was approximately 330bp (both from primers

F2 and R3). The two products were both present in *C. destructor* 30% and 50% embryo RNA. Figure 3.6 also shows the results of the first round (primers F1 and R1) RT-PCR of RNA from 30% and 50% *C. destructor* embryos. A multitude of bands are observed, constituting much non-specific background. Raising the stringency of this reaction eventually produced a single band. This product was purified and sequenced, and was found not to be a *pgIII* gene. Thus a nested technique was certainly required.

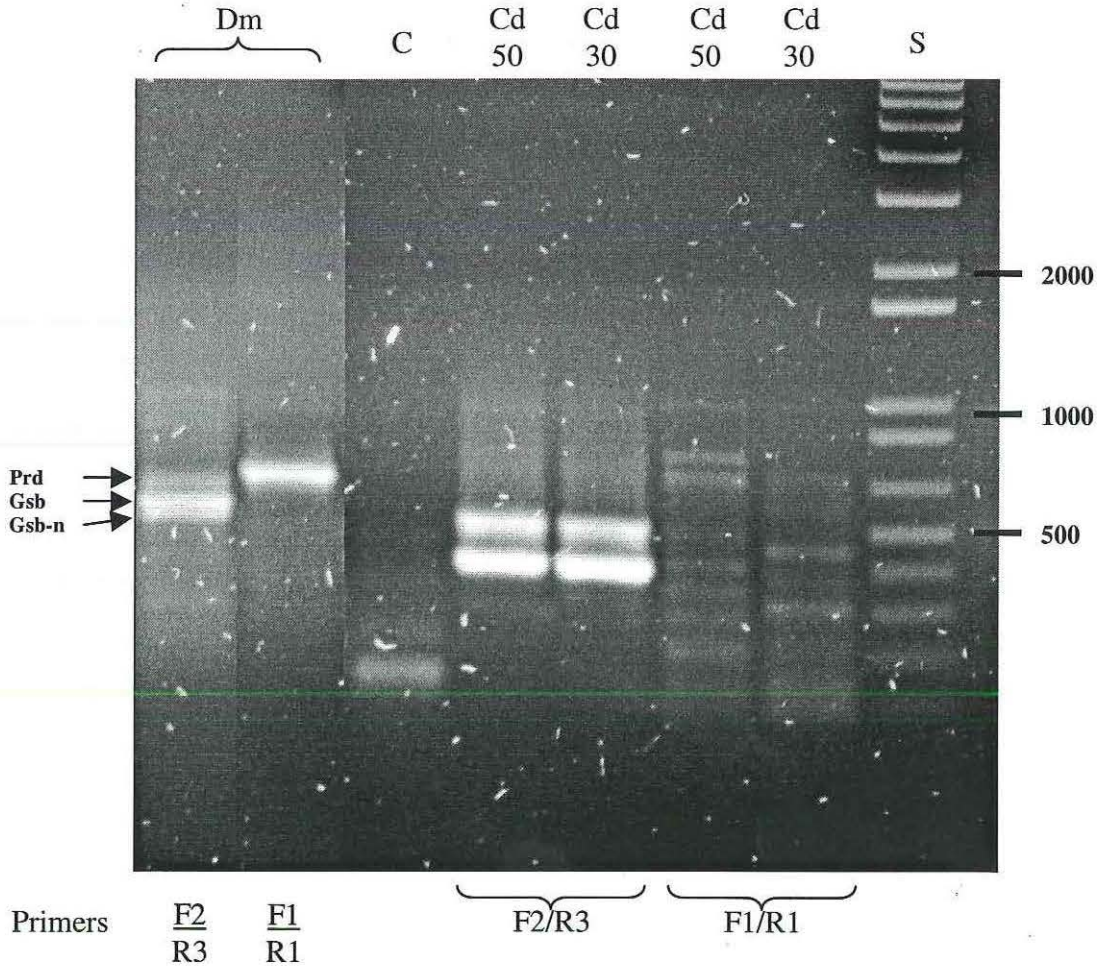


Figure 3.6: Ethidium bromide stained agarose gel of first round RT-PCR (primers F1/R1) and nested RT-PCR (primers F2/R3) of RNA isolated from *C. destructor* embryos at 30% (Cd 30) and 50% (Cd 50) of development and *D. melanogaster* mixed stage 8 and stage 14 embryos (Dm) showing products corresponding to the predicted sizes of *Prd*, *Gsb*, and *Gsb-n* (as indicated). DNA size standard (S) and blank (no nucleic acid template) control (C) are also shown.

The predicted amplified regions for other invertebrate *pgIII* genes using primers F2 and R3 are; PaxD: 614bp; Prd: 490bp; Tupax3/7: 472bp; Gsb: 430bp; Pby1: 427bp; Pby2: 421bp; Gsb-n: 418bp (Figure 2.3) (Dearden *et al.*, 2002; Davis *et al.*, 2001; Miller *et al.*, 2000; Baumgartner *et al.*, 1987; Kilcherr *et al.*, 1986). So it may be predicted that *pgIII* genes in *C. destructor* will produce amplification products between approximately 400bp and 700bp using primers F2 and R3. Thus the larger product (420bp) fits within this predicted range, however the smaller product (330bp) does not. Raising the stringency conditions of nested RT-PCR reactions did not eliminate this smaller product.

The optimal conditions for this amplification were the same as for nested RT-PCR *D. melanogaster* embryo *pgIII* amplification (section 3.4.3). Lowering the stringency did not result in an increase in amplification products within the predicted size range for pgIII genes (400-700bp), although spurious background did result. Nested RT-PCR using primers DF1/DR1 (first round) followed by DF2/DR2 (second round) (Davis *et al.*, 2001) did not result in the amplification of any other products from *C. destructor* 30% or 50% embryo RNA.

3.6 Sequencing of *Cherax destructor* cDNAs

The presence of a *pgIII* gene in *C. destructor* was confirmed by sequencing. Nested RT-PCR reaction products from *C. destructor* embryo (at 50% of development) RNA using primers F1/R1 (first round) and F2/R2 (second round) were gel purified and sequenced using BigDye terminator chemistry. Nested RT-PCR using primers F1/R1 (first round) and F2/R3 (second round) generated two cDNA products. The small cDNA product is 331bp and the large cDNA product is 424bp. The large and small cDNA products were each sequenced seven times using primers F2 and R3 (three forward and four reverse). A single *pgIII* gene from *C. destructor* was identified by these sequences (as described in section 3.8). I have designated this gene *Cdpax3/7* (*Cherax destructor* paired box gene similar to *Pax3* and *Pax 7*). The use of any combination of different primer pairs (including DF1/DR1 and DF2/DR2; Davis *et al.*, 2001) did not result in the identification of any other *pgIII* genes in RNA isolated from *C. destructor*.

To obtain a larger length of sequence, nested RT-PCR reactions were conducted on *C. destructor* embryo (50%) RNA using primers DF1/DR1 (first round) and primers F1/R1 (second round). The products of this reaction were gel purified and forward sequenced using primer F1, and reverse sequenced using primer R1. This technique facilitated the sequencing of a 514bp region of the larger cDNA and a 421bp region of the smaller cDNA (Figure 3.8).

3.6.1 *Cdpax3/7* Is Expressed as Two Alternate Transcripts

Sequence analysis of both the large (424bp) and small (331bp) nested RT-PCR products from *C. destructor* indicates that both cDNAs are products of the same gene. Sequences were confirmed in both forward and reverse orientation, and no differences were detected other than a large section missing from the smaller (331bp) cDNA. *Cdpax3/7* is expressed as two alternate transcripts, the larger (424bp) product designated *Cdpax3/7-a* and the smaller (331bp) product designated *Cdpax3/7-b*. The *Cdpax3/7-b* mRNA is a shortened alternate transcript in which a 93 nucleotide section has been spliced from the original transcript (as indicated by the sequence of *Cdpax3/7-a*).

3.7 Expression of *Cdpax3/7*

Nested RT-PCR using primers F1/R1 (first round) and F2/R3 (second round) was then used to assay *Cdpax3/7* expression during embryogenesis and adult muscle regeneration. *Cdpax3/7* expression was assayed in RNA isolated from *C. destructor* whole embryos at 30% and 50% of development, moult cycle regenerating claw closer muscle (premoult, postmoult, and intermoult), and epimorphically regenerating cheliped at the papilla and closed dactyl stages where muscle regeneration occurs.

3.7.1 *Cdpax3/7* Is Expressed During Embryonic Development

Cdpax3/7 is expressed during naupliar (30% of development) and postnaupliar (50% of development) embryonic development in *C. destructor*. Nested RT-PCR assays of RNA isolated from *C. destructor* whole embryos at 30% and 50% of development both generated 424bp and 331bp cDNAs (Figure 3.7) indicating expression of *Cdpax3/7*. Both transcripts

(*Cdpax3/7-a* and *Cdpax3/7-b*) were expressed at 30% and 50% of embryonic development (Figure 3.7). The expression of *Cdpax3/7* at 50% of development coincides with the first external signs of limb bud (cheliped) development (Figure 3.2).

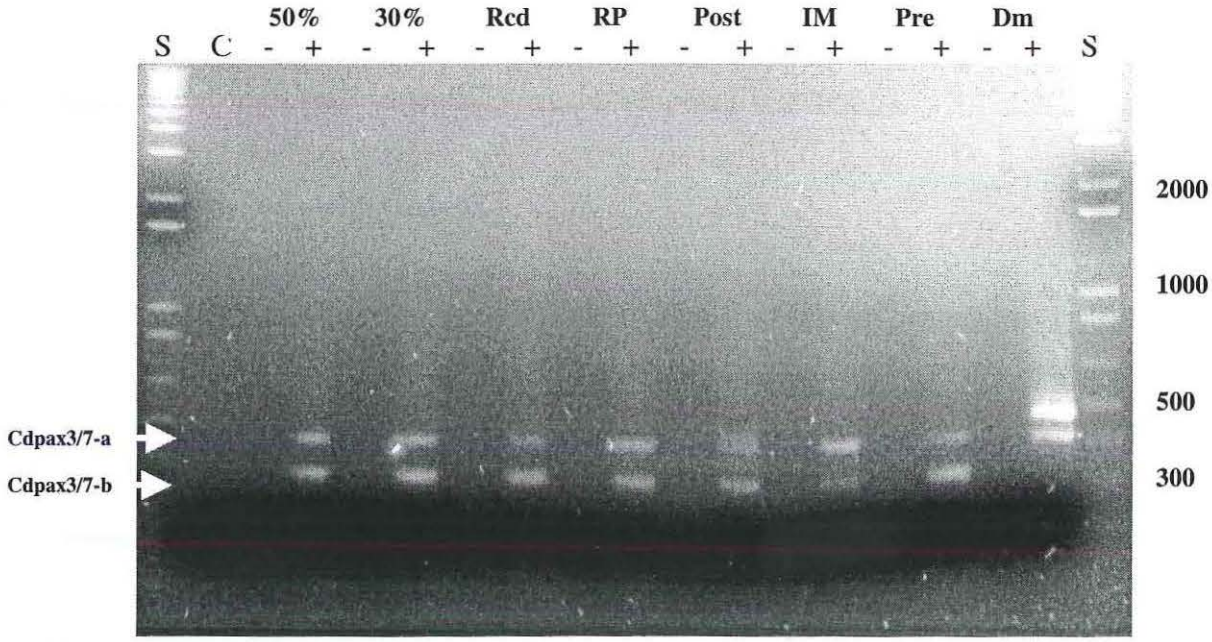


Figure 3.7: Ethidium bromide stained agarose gel showing nested RT-PCR assays of *C. destructor* moult cycle regenerating claw closer muscle samples at premoult (Pre), intermoult (IM), and postmoult (Post), epimorphically regenerating cheliped at papilla (RP) and closed dactyl stages (Rcd), and embryos at 30% (30%) and 50% (50%) of development showing expression of both alternate transcripts of *Cdpax3/7* (*Cdpax3/7-a* and *Cdpax3/7-b*; as indicated). Reactions performed with reverse transcriptase enzyme (experimental reactions) indicated by (+), control reactions preformed without reverse transcriptase enzyme (expression control reactions) indicated by (-). Also shown are *D. melanogaster* embryos (Dm; mixed stage 8 and stage 14), DNA size standard (S) and blank (no nucleic acid template) control (C).

3.7.2 *Cdpax3/7* Is Expressed During Adult Muscle Regeneration

Cdpax3/7 is expressed during moult cycle muscle regeneration. Nested RT-PCR assays of RNA isolated from premoult, intermoult, and postmoult claw closer muscle all generate both 424bp and 331bp cDNAs (Figure 3.7). This indicates that *Cdpax3/7* is expressed in intermoult claw closer muscle and claw closer muscle undergoing moult-cycle regeneration (both late-premoult and postmoult). Both alternate transcripts were detected in all moult cycle regeneration muscle samples (Figure 3.7).

Cdpax3/7 is also expressed during epimorphic cheliped regeneration. Nested RT-PCR assays of RNA isolated from epimorphically regenerating cheliped tissue at the papilla as well as closed dactyl stages generated both 424bp and 331bp cDNAs (Figure 3.7). This indicates th' . *Cdpax3/7* is expressed in the epimorphically regenerating cheliped both before (papilla stage) and after (closed dactyl stage) muscle differentiation (Cutler *et al.*, 2002). Both alternate transcripts, *Cdpax3/7-a* (424bp) and *Cdpax3/7-b* (331bp) were detected in papilla and closed-dactyl regenerating limb (Figure 3.7).

3.8 *Cdpax3/7* Sequence Analysis

Sequence data obtained for *C. destructor Cdpax3/7* cDNA sequences was analysed for homology/similarity to other known *Pax* genes. Due to the functional significance of Pax proteins, the protein sequence of the *Cdpax3/7* cDNA sequences was deduced, and this was used for comparisons to other Pax proteins. Structural elements of the Cdpax3/7 protein were then deduced by comparison to the known structure of other Pax proteins. The phylogenetic relationship of Cdpax3/7 to other known Pax proteins was then inferred by maximum likelihood analysis.

3.8.1 *Cdpax3/7* Database Search and Translation

A translated database search (BLASTx) revealed high homology of the deduced protein sequence of the *Cdpax3/7* cDNA sequence to known pgIII proteins in GenBank. This translated search was used to deduce an open reading frame and the corresponding translation of the Cdpax3/7 protein, by reference to the Pby1 protein (the most similar protein identified by BLASTx). The deduced protein sequence of Cdpax3/7 is given in Figure 3.8.

BLASTx revealed that the deduced protein sequence of Cdpax3/7 is most similar to the two pgIII proteins isolated from the grasshopper *S. americana* Pby1 (81% identity) and Pby2 (75% identity) and fruit fly *D. melanogaster* Gsb-n (77% identity). The deduced protein sequence of Cdpax3/7 is also highly similar to other protostome and deuterostome pgIII protein sequences (Table 3.3).

(A)

```
TTG CGT GTT TCT CAT GGC TGC GTC TCC AAG ATC CTC AAC CGC TAC CAG GAG ACC GGC TCC   60
 L   R   V   S   H   G   C   V   S   K   I   L   N   R   Y   Q   E   T   G   S   20

ATC AGG CCG GGC GTC ATC GGC GGC TCC AAG CCC AAA GTA ACC ACG CCG GAT ATT GAG AAG  120
 I   R   P   G   V   I   G   G   S   K   P   K   V   T   T   P   D   I   E   K   40

CGC ATC TAC GAC TAC AAG AAG GAG AAT CCG GGC ATA TTT TCT TGG GAG ATC AGA GAC GAC  180
 R   I   Y   D   Y   K   K   E   N   P   G   I   F   S   W   E   I   R   D   D   60

ACC ATC AAG GAG GGT GTG GTG GAC AAG GCG TCA GCT CCC TCC GTC AGC TCC ATC AGC AGG  240
 T   I   K   E   G   V   V   D   K   A   S   A   P   S   V   S   S   I   S   R   80

ATC CTC AGA GGT GGT AAG AGG GAT GAC GAC CCT CGC AAA GAC CAC AGC ATC GAC GGC ATC  300
 I   L   R   G   G   K   R   D   D   D   P   R   K   D   H   S   I   D   G   I  100

CTC GGA GGT GGA GGC AGT GAC GAG AGT GAC ATC GAG AGT GAG CCG GGG ATC CCC CTC AAG  360
 L   G   G   G   G   S   D   E   S   D   I   E   S   E   P   G   I   P   L   K  120

CGT AAG CAG AGG CGC TCG AGG ACC ACC TTC ACT GCA GAG CAG CTG GAG GTG CTG GAG CGC  420
 R   K   Q   R   R   S   R   T   T   F   T   A   E   Q   L   E   V   L   E   R  140

TCC TTC GAG AAG ACA CAG TAC CCA GAC GTC TAC ACC AGG GAG GAA CTG GCT CAG AAA GCT  480
 S   F   E   K   T   Q   Y   P   D   V   Y   T   R   E   E   L   A   Q   K   A  160

AAG CTG ACG GAG GCC CGN ATA CAA GTG TGG                                         510
 K   L   T   E   A   R   I   Q   V   W                                         170
```

(B)

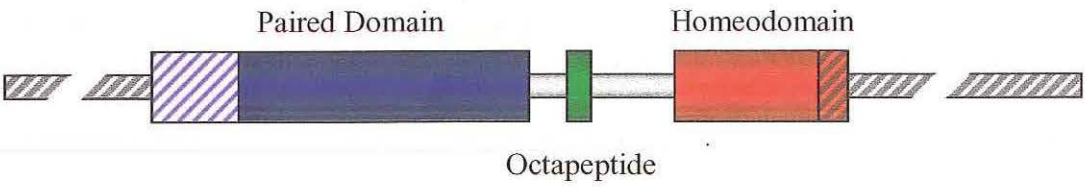


Figure 3.8: Molecular characterisation of *Cdpax3/7*. (A) Nucleotide and deduced amino acid sequence of partial *Cdpax3/7* cDNA; boxed sequence corresponds to the conserved paired box/paired domain (blue), homeobox/homeodomain (red) and octapeptide (green) sequence. (B) Schematic representation of Cdpax3/7 protein structure; hashing represents predicted regions (not sequenced); gaps in sections indicate lack of certainty about length.

Table 3.3: Amino acid identity of Cdpax3/7 protein sequence to other pgIII sequences (total, partial paired domain [PD], partial homeodomain [HD], and partial PD plus partial HD).

Protein	Overall identity	PD identity	HD identity	PD+HD identity
Pby1	81%	84%	81%	83%
Gsb-n	77%	77%	85%	80%
Pby2	75%	81%	75%	79%
Amphipax3/7	73%	72%	81%	75%
Tupax3/7	70%	76%	72%	74%
Pax7	68%	72%	83%	76%
Pax3	67%	76%	85%	79%
Gsb	67%	77%	73%	76%
Prd	63%	73%	77%	75%
PaxD	41%	55%	73%	62%

3.8.2 Protein Structure

A 170 amino acid protein sequence was deduced from the *Cdpax3/7* cDNA sequence. Comparisons with other pgIII proteins (Prd, Pax3 and PAX6) reveal that Cdpax3/7 contains a paired domain (PD) and a homeodomain (HD) separated by a PD-HD linker region containing the conserved octapeptide (HSIDGILG) motif present in the majority of pgIII proteins (Balczarek *et al.*, 1997). Sequence was obtained from the 43rd amino acid of the 128 amino acid PD through to the 48th amino acid of the 60 amino acid HD, including a linker region of 37 amino acids (including the octapeptide) separating the two DNA binding domains (Figure 3.8).

Because of the functional significance of the PD and HD DNA-binding domains, structural conservation within these regions is of greater functional significance than that of the linking regions. This is indicated by analysis of the partial PD and partial HD of Cdpax3/7 relative to other pgIII proteins (Table 3.3). All known pgIII proteins are homologous to Cdpax3/7 within the PD and HD regions, displaying between 2% and 21% greater homology than the linker regions (Table 3.3), indicating that PD nd HD sequences are diverging at a low rate.

Moreover, the amino acid identity within binding regions of Cdpax3/7 and Pax3/Pax7 is remarkably homologous (79% and 76% identity, respectively), on a par with the identity to some of the arthropod pgIII proteins from animals which are much more closely related to *C. destructor*. In all of these analyses, Cdpax3/7 exhibits greatest amino acid identity with Pby1, Pby2 and Gsb-n.

3.9 Phylogenetic Analysis of Cdpax3/7

Maximum likelihood analysis was performed on Cdpax3/7 and other representative Pax proteins to infer the phylogenetic relationship of Cdpax3/7 within the family of known Pax proteins and to assign Cdpax3/7 to a specific Pax group. PD and HD DNA binding domains show a greater degree of conservation than other sections of pgIII proteins (Breitling & Gerber, 2000; Groger *et al.*, 2000; Sun *et al.*, 1997), and so are believed to give a more accurate reflection of functionally important relationships between orthologous

and paralogous genes (Groger *et al.*, 2000; Agosti *et al.*, 1996; Felsenstein & Sober, 1986). Thus phylogenetic analysis was performed using PD and HD amino acid sequences. As the full 128 amino acid PD and 60 amino acid HD sequences were not obtained for Cdpax3/7, this analysis was conducted only on partial PD (86 amino acids) and partial HD (48 amino acids) amino acid sequences.

Maximum likelihood analysis was chosen for the analysis of the phylogenetic relationships of Cdpax3/7 within the Pax family, instead of other methods such as parsimony or neighbour-joining. Maximum likelihood analysis facilitates the calculation of phylogram branch lengths that are proportional to the actual amino acid change between corresponding proteins, a feature that is not available in the two other methods of phylogenetic analysis (Freeman & Herron, 2004). Maximum likelihood phylograms thus infer evolutionary relationships in their branching pattern and give a relative measure of evolutionary distance in their branch lengths (the longer the branch between two proteins, the more divergent their amino acid sequence). The branching pattern of the Pax family maximum likelihood phylogram was statistically analysed by bootstrapping, with 100 pseudoreplicates generated. Bootstrap values less than 40% were considered to be significantly uncertain in this analysis, a probability value that has been used previously for the phylogenetic analysis of Pax proteins (Miller *et al.*, 2000; Groger *et al.*, 2000).

The deduced protein sequence for the partial PD and partial HD of Cdpax3/7 and of other representative Pax proteins (Pax1 to 9 from *M. musculus*; Pox-meso, Pox-neuro, Eyeless, Sparkling, Prd, Gsb, and Gsb-n from *D. melanogaster*; Pby1 and Pby2 from *S. americana*; Tupax3/7 from *T. urticae*; Amphipax1, 2, 3/7, and 6 from *B. floridae*; HrPax258 from *H. roretzi*; and PaxA, B, C, and D from *A. millepora*; Table 2.2) were aligned using ClustalW (Appendix). Maximum likelihood analysis was performed using the Jones-Taylor-Thornton (1992) model of amino acid change for branch length estimation. The resultant phylogram is presented in Figure 3.9.

3.9.1 Cdpax3/7 Is a Member of Pax Group III

Maximum likelihood analysis of the partial PD and partial HD indicate that Cdpax3/7 is unequivocally a pgIII protein. In the maximum likelihood phylogram (Figure 3.9) all pgIII

proteins are grouped together as a distinct cluster with high bootstrap support (90%). Cdpax3/7 is positioned within this pgIII cluster, and this positioning is statistically supported with a reasonable bootstrap support value (54%; Figure 3.9).

The pgIII cluster is divided into two clades representing pgIII proteins from deuterostomes (*M. musculus* Pax3 and Pax7, and *B. floridae* Amphipax3/7) with a bootstrap support of 61%, and pgIII proteins from arthropods (*T. urticae* Tupax3/7; *S. americana* Pby1 and Pby2; *D. melanogaster* Prd, Gsb, and Gsb-n) with a bootstrap support of 81% in the maximum likelihood phylogram (Figure 3.9). Cdpax3/7 is included in a well distinguished clade with other arthropod pgIII proteins, to the exclusion of deuterostome pgIII proteins. These two clades are separated by the branching of the PaxD protein from the Cnidarian *A. millepora*, the placement of which is not statistically resolved (given a bootstrap support value of 35%).

3.9.2 The Phylogeny of Cdpax3/7

The maximum likelihood phylogram (Figure 3.9) branching pattern leading to Cdpax3/7 infers the phylogenetic origins of the Cdpax3/7. The centre of the phylogram represents an ancestral Pax protein. This protein gives rise to two Pax proteins (the progenitors of supergroups I and II, first defined by Sun *et al.*, 1997), the ancestor of supergroup II then gives rise to the progenitors of groups I and III. This progression is given 69% bootstrap support and is in agreement with previous analyses (Miller *et al.*, 2000; Balczarek *et al.*, 1997). The monophyletic origin of pgIII is strongly supported (90%) and pgIII is still monophyletic at the divergence of Protostomia and Deuterostomia. Both Tupax3/7 and Cdpax3/7 branch from a point in the pgIII phylogeny where only a single protostome pgIII protein is present, branching patterns that are statistically supported (81% and 54%, respectively). This branching pattern indicates that both the nearest common ancestor of Crustacea and Insecta, and the nearest common ancestor of Chelicerata and Crustacea, probably had only a single pgIII protein. The phylogenetic relationships of the insect pgIII proteins are not statistically resolved by this analysis. However, the branching pattern does indicate that Cdpax3/7 is positioned in an ancestral location to all three *D. melanogaster* and two *S. americana* pgIII proteins, with reasonable bootstrap support (54%).

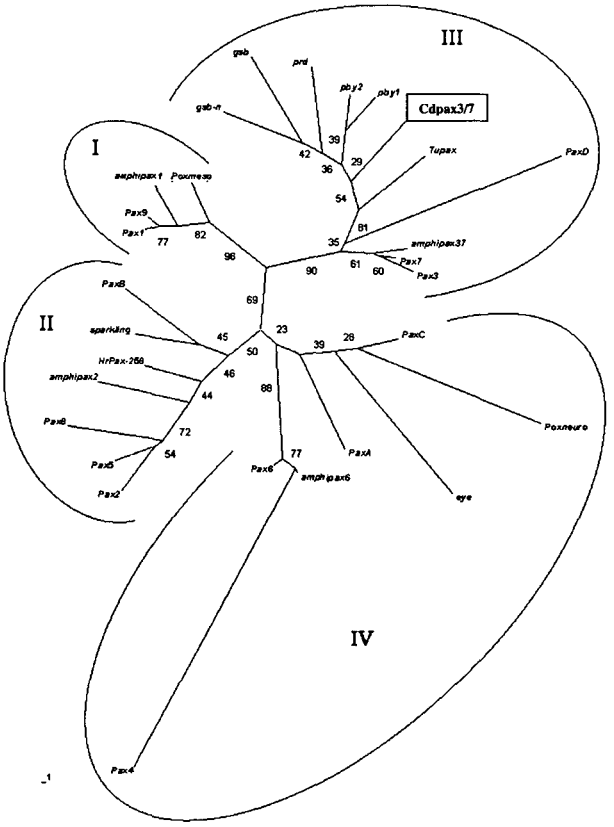


Figure 3.9: Unrooted maximum likelihood phylogram of representative Pax protein sequences. Branch lengths are given as the expected fraction of amino acid change using the Jones-Taylor-Thornton model (scale is 1.0 = 100 PAM values [percent of accepted mutations]; corresponding to 1 amino acid mutation per 100 amino acids; Jones *et al.*, 1992). Numbers shown are bootstrap percentages of 100 pseudoreplicates. Sequences are: Cdpax3/7 from *Cherax destructor*; Tupax3/7 from *Tetranychus urticae*; Pby1 and Pby2 from *Schistocerca americana*, Pax 1-9 from *Mus musculus*; PaxA, B, C, and D from

Acropora millepora; amphipax 1, 2, 3, and 6 from *Branchiostoma floridae*; Prd, Gsb, Gsb-n, Eyeless (eye), Poxneuro, Poxmeso, and Sparkling from *Drosophila melanogaster*. Pax groups I, II, III, and IV are those originally defined by Balczarek *et al.* (1997).

The relative branch lengths leading to each of the arthropod pgIII proteins are an indicator of the divergence of their amino acid sequences (Felsenstein & Churchill, 1996). The branch length leading to Cdpax3/7 is of comparable length to the branches leading to the majority of the other arthropod pgIII proteins. Of note is the shorter branches leading to Pby1 and Pby2, indicating that the sequences of these two proteins are less divergent from one another than they are from the other arthropod members of pgIII. Also of interest are the relative lengths of the branches leading to the deuterostome pgIII proteins compared to the lengths of branches leading to the arthropod pgIII proteins. The branches separating Pax3, Pax7 and Amphipax3/7 are substantially shorter than branches in the arthropod pgIII clade. This indicates that the relative amino acid divergence of pgIII proteins within the arthropods is greater than in the chordates. Also, all individual arthropod pgIII proteins are further from the point of protostome/deuterostome divergence than the chordate proteins, indicating that the protostome pgIII proteins have diverged at a greater rate than the deuterostome pgIII proteins.

Because a suitable predecessor protein to all Pax proteins has not yet been found (Breitling & Gerber, 2000; Miller *et al.*, 2000), the maximum likelihood phylogram of the entire Pax family does not imply a root (Figure 3.9). PaxD from the cnidarian *A. millepora* is the most basal pgIII protein isolated to date, and is ancestral to all protostome and deuterostome pgIII proteins (Miller *et al.*, 2000), thus it may be reliably used as the root of a pgIII phylogram (Dearden *et al.*, 2002; Davis *et al.*, 2001). Thus a maximum likelihood phylogram was constructed from pgIII protein sequences with PaxD as the root of the phylogram, and is presented in Figure 3.10. The topology of this phylogram is identical to the phylogram for the full Pax family (Figure 3.9), indicating that the topology of pgIII proteins presented here is robust whether assessed in relation to other Pax proteins or assessed in relation only to pgIII proteins.

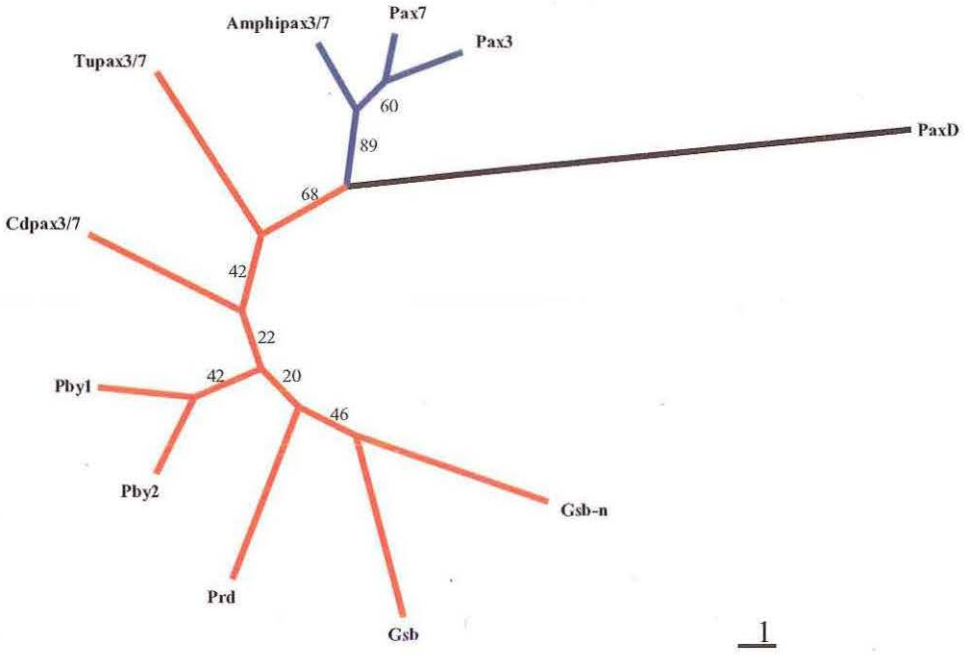


Figure 3.10: Maximum likelihood phylogram of representative Pax group III protein sequences rooted with PaxD from the Cnidarian *Acropora millepora*. Branch lengths are given as the expected fraction of amino acid change using the Jones-Taylor-Thornton model (scale is 1.0 = 100 PAM values [percent of accepted mutations]; corresponding to 1 amino acid substitution per 100 amino acids; Jones *et al.*, 1992). Numbers shown are bootstrap percentages of 100 pseudoreplicates. Sequences are: Cdpax3/7 from *Cherax destructor*; Tupax3/7 from *Tetranychus urticae*; Pby1 and Pby2 from *Schistocerca americana*, Pax3 and Pax7 from *Mus musculus*; Pax D from *A. millepora*; Amphipax3/7 from *Branchiostoma floridae*; Prd, Gsb, and Gsb-n from *Drosophila melanogaster*. Root of phylogram is represented by a black line, deuterostome lineage is represented by blue lines, and the protostome lineage is represented by red lines.

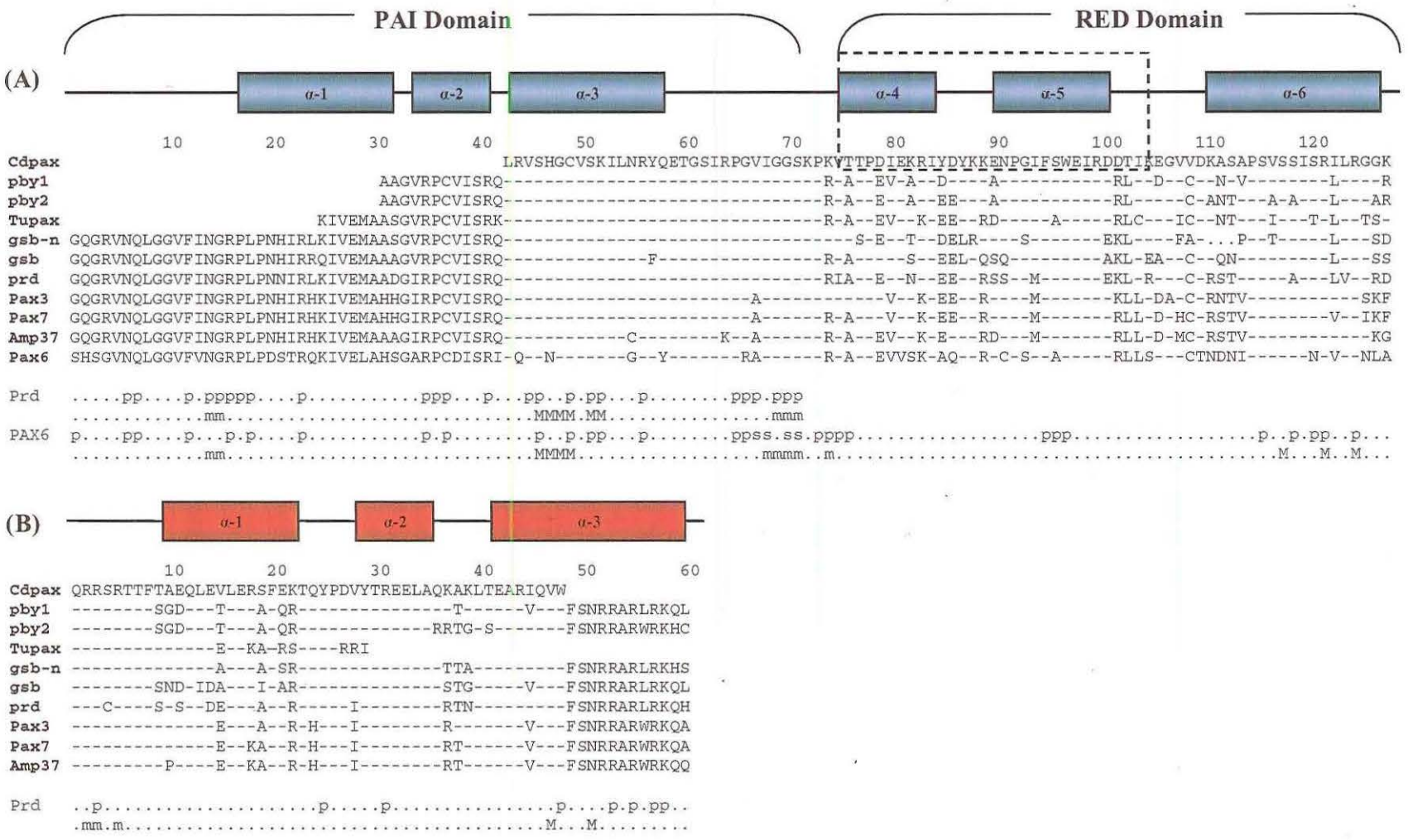


Figure 3.11: Alignment of Cdpax3/7 against representative (see Methods for accession numbers) Pax protein (A) PD and (B) HD sequences showing α-helical structure and PAI and RED domains of the PD (above alignment) and DNA contacting positions from crystal structures of PDs of Prd (Xu *et al.*, 1995) and PAX6 (Xu *et al.*, 1999), and HD of Prd (Wilson *et al.*, 1995) (below alignment). DNA contacts are; p: phosphate; s: sugar: m: minor groove; M: major groove. Dashes (-) indicate amino acids identical to Cdpax3/7, dots (.) indicate no amino acid, gaps ( ) indicate regions not yet sequenced, Dashed box (---) region indicates the section missing from Cdpax3/7-b alternate isoform.

3.10 DNA-binding domains

The crystal structures of two Pax proteins (*D. melanogaster* Prd and *Homo sapiens* PAX6) bound to DNA sequences have been solved. Both proteins have been found to make direct contact with DNA only with amino acids within the PD and HD DNA binding domains (Xu *et al.*, 1995; Wilson *et al.*, 1995; Xu *et al*,. 1999). Amino acids at certain positions within the PD and HD of Prd and PAX6 make direct contact with DNA (Xu *et al.*, 1995; Wilson *et al.*, 1995; Xu *et al*,. 1999). Amino acid substitutions at these DNA contacting positions in the *M. musculus* Pax3 protein have been shown to exert an overriding effect on the DNA binding properties of the protein (Apuzzo & Gros, 2002), and are of greater functional significance to a Pax protein than the overall amino acid composition of the protein. A further assessment of Cdpax3/7 relative to other Pax proteins was made via analysis of PD and HD sequences and their respective putative DNA binding amino acid sites.

3.10.1 Analysis of the Cdpax3/7 Paired Domain

The paired domain (PD) is a bipartite DNA binding domain consisting of two helix-turn-helix motifs. The N-terminal (PAI domain) consists of three α-helices (helix 1, 2, and 3) as does the C-terminal (RED domain) (helix 3, 4, and 5). The crystal structures of Prd and PAX6 indicate that these two Pax proteins differ in their use of the RED domain in DNA binding. The PD of *H. sapiens* PAX6 protein binds using both its PAI and RED domains (Xu *et al.*, 1999), whilst the PD of *D. melanogaster* Prd only utilises its PAI domain, with its RED domain not making any DNA contact (Xu *et al.*, 1995).

Analysis of the crystal structure of the *D. melanogaster* Prd protein PD bound to a DNA recognition sequence has revealed that Prd makes contact with the DNA phosphate backbone using amino acids at positions 6, 7, 12, 14, 15, 16, 17, 18, 23, 35, 36, 37, 41, 45, 46, 49, 51, 52, 56, 65, 66, 67, 69, 70, and 71 of the PD (Xu *et al.*, 1995) (Figure 3.11). Prd contacts DNA in the minor groove using amino acids at positions 14, 15, 69, 70, and 71 of the PD, and contacts the major groove with amino acids at positions 46, 47, 48, 49, 51, and 52 of the PD (α-helix 3 of the PD, designated the recognition helix) (Xu *et al.*, 1995) (Figure 3.11).

All amino acids of the Cdpax3/7 PD at positions corresponding to DNA contacting positions in the PD of the Prd protein are conserved between Cdpax3/7 and Prd (Figure 3.11). Also, the PD of Cdpax3/7 differs from the three chordate pgIII proteins PDs, mouse Pax3 and Pax 7 and amphioxus Amphipax3/7 (amp37) in only a single DNA contact position (amino acid 67); where Cdpax3/7 contains a valine residue, the chordate proteins contain an alanine residue. Thus my results indicate that Cdpax3/7 may be capable of utilising its PD to bind DNA in a similar manner to Prd, and even Pax3 and Pax7 proteins.

The binding pattern of Prd may differ from binding patterns of other pgIII proteins, insofar as no data exists for DNA binding by the Prd RED domain. The RED domain of the Prd PD does not contact DNA (Xu *et al.*, 1995) and is dispensable for normal *in vivo* function (Bertuccioli *et al.*, 1996), whereas the RED domain of Pax3 and Pax7 does influence protein function/DNA binding (Vogan *et al.*, 1996). The crystal structure of PAX6 bound to its DNA recognition sequence contains DNA binding sites (Xu *et al.*, 1999). It has been demonstrated that the murine Pax3 RED domain recognises identical DNA sequences to the PAX6 RED domain (Vogan & Gros, 1997), indicating that the binding sites of the PAX6 RED domain are applicable to pgIII proteins. Therefore, the DNA binding positions of PAX6 are compared with Cdpax3/7 and other pgIII proteins, to compare DNA-binding sites for the RED domain of the PD (Figure 3.11).

All amino acids of the RED domain of the PD that contact the DNA major groove are conserved between Cdpax3/7, PAX6 and all pgIII proteins. However amino acids at positions 74, 76, 96 and 121 differ between PAX6 and Cdpax3/7 proteins. These amino acids are known to contact DNA via the phosphate backbone or minor groove in the PAX6 protein (Xu *et al.*, 1999). Amino acid position 119 is also of some relevance and has been shown to contact the DNA phosphate backbone in PAX6. Prd and Pby2 contain an alanine residue, whereas PAX6, Cdpax3/7 and the chordate proteins contain a serine residue.

3.10.2 Analysis of the Cdpax3/7 Homeodomain

Crystal structure analysis of the HD of *D. melanogaster* Prd protein bound to its DNA recognition sequence indicates that the Prd HD makes DNA phosphate contact using amino acids at positions 3, 25, 31, 48, 53, 55, 57, and 58, and contacts the minor groove using

amino acids at positions 2, 3, and 5, and contacts the major groove using amino acids at positions 47 and 51 (α-helix 3 of the HD, the recognition helix) (Wilson *et al.*, 1995) (Figure 3.11). All amino acids at DNA contacting positions in the HD of the Prd protein are conserved within the sequenced region (amino acids 1 to 48) of the HD of Cdpax3/7 (Figure 3.11). Remarkably, all of the amino acids of the HD that contact DNA are conserved between all pgIII proteins (Figure 3.11).

Therefore, this analysis indicates that the PD and HD of the protein sequence deduced from the *Cdpax3/7* cDNA isolated from *C. destructor* have remarkably high homology, both overall and in specific DNA contacting amino acid positions, to *D. melanogaster* Prd, and importantly to vertebrate Pax3 and Pax7.

3.11 Cdpax3/7-b and DNA Binding Function

The *Cdpax3/7-b* cDNA is a shortened alternate transcript in which a 93 nucleotide section has been spliced out of the original transcript (indicated by comparison to the sequence of *Cdpax3/7-a*). Structural analysis of the deduced protein sequence for the alternate Cdpax3/7-b isoform indicates that the spliced section lacks 31 amino acids and causes the exclusion of the fourth and fifth α-helices of the PD (amino acid residues 75 to 105) from the final alternate Cdpax3/7-b isoform (Figure 3.12), indicating that splice sites for the generation of the *Cdpax3/7-b* alternate transcript correspond to amino acid positions 75 and 105 of the PD (Figure 3.12).

To assess the putative DNA binding significance of the alternate Cdpax3/7-b isoform, the deduced amino acid sequence has been compared with the crystal structure deduced DNA binding sites of Prd and PAX6 proteins (Figure 3.11). The 31 amino acid section spliced from Cdpax3/7-b corresponds to a region of Prd protein that is not used for DNA binding (Xu *et al.*, 1995). This region in the Pax6 protein contacts DNA using 6 different amino acids (making phosphate or minor groove contacts), but does not make any DNA Major groove contacts in this region (DNA sequence recognition contacts) (Xu *et al.*, 1999). Comparative analysis of pgIII proteins indicates that the spliced region corresponds to a region that is also much less conserved than the section of pgIII proteins immediately N-terminal to it (the PAI domain) (Figure 3.11). Therefore, these results demonstrate that the

alternate Cdpax3/7-b isoform lacks the first two α-helices of the RED domain of the PD, and therefore all of the DNA contacting positions contained within them.

(A)

```
a  TTGCGTGTTTCTCATGGCTGCGTCTCCAAGATCCTCAACCGCTACCAGGAGACCGGCTCCATCAGGCCGGGCGTC   75
a   L  R  V  S  H  G  C  V  S  K  I  L  N  R  Y  Q  E  T  G  S  I  R  P  G  V   25
b  TTGCGTGTTTCTCATGGCTGCGTCTCCAAGATCCTCAACCGCTACCAGGAGACCGGCTCCATCAGGCCGGGCGTC   75
b   L  R  V  S  H  G  C  V  S  K  I  L  N  R  Y  Q  E  T  G  S  I  R  P  G  V   25

a  ATCGGCGGCTCCAAGCCCAAAGTAACCACGCCGGATATTGAGAAGCGCATCTACGACTACAAGAAGGAGAATCCG   150
a   I  G  G  S  K  P  K  V  T  T  P  D  I  E  K  R  I  Y  D  Y  K  K  E  N  P   50
b  ATCGGCGGCTCCAAGCCCAAA                                                         96
b   I  G  G  S  K  P  K                                                          32

a  GGCATATTTTCTTGGGAGATCAGAGACGACACCATCAAGGAGGGTGTGGTGGACAAGGCGTCAGCTCCCTCCGTC   225
a   G  I  F  S  W  E  I  R  D  D  T  I  K  E  G  V  V  D  K  A  S  A  P  S  V   75
b                                          GAGGGTGTGGTGGACAAGGCGTCAGCTCCCTCCGTC   132
b                                           E  G  V  V  D  K  A  S  A  P  S  V   44

a  AGCTCCATCAGCAGGATCCTCAGAGGTGGTAAGAGGGATGACGACCCTCGCAAAGACCACAGCATCGACGGCATC   300
a   S  S  I  S  R  I  L  R  G  G  K  R  D  D  D  P  R  K  D  H  S  I  D  G  I   100
b  AGCTCCATCAGCAGGATCCTCAGAGGTGGTAAGAGGGATGACGACCCTCGCAAAGACCACAGCATCGACGGCATC   207
b   S  S  I  S  R  I  L  R  G  G  K  R  D  D  D  P  R  K  D  H  S  I  D  G  I   69

a  CTCGGAGGTGGAGGCAGTGACGAGAGTGACATCGAGAGTGAGCCGGGGATCCCCCTCAAGCGTAAGCAGAGGCGC   375
a   L  G  G  G  G  S  D  E  S  D  I  E  S  E  P  G  I  P  L  K  R  K  Q  R  R   125
b  CTCGGAGGTGGAGGCAGTGACGAGAGTGACATCGAGAGTGAGCCGGGGATCCCCCTCAAGCGTAAGCAGAGGCGC   282
b   L  G  G  G  G  S  D  E  S  D  I  E  S  E  P  G  I  P  L  K  R  K  Q  R  R   94

a  TCGAGGACCACCTTCACTGCAGAGCAGCTGGAGGTGCTGGAGCGCTCCTTCGAGAAGACACAGTACCCAGACGTC   450
a   S  R  T  T  F  T  A  E  Q  L  E  V  L  E  R  S  F  E  K  T  Q  Y  P  D  V   150
b  TCGAGGACCACCTTCACTGCAGAGCAGCTGGAGGTGCTGGAGCGCTCCTTCGAGAAGACACAGTACCCAGACGTC   357
b   S  R  T  T  F  T  A  E  Q  L  E  V  L  E  R  S  F  E  K  T  Q  Y  P  D  V   119

a  TACACCAGGGAGGAACTGGCTCAGAAAGCTAAGCTGACGGAGGCCCGNATACAAGTGTGG                  510
a   Y  T  R  E  E  L  A  Q  K  A  K  L  T  E  A  R  I  Q  V  W                   170
b  TACACCAGGGAGGAACTGGCTCAGAAAGCTAAGCTGACGGAGGCCCGNATACAAGTGTGG                  417
b   Y  T  R  E  E  L  A  Q  K  A  K  L  T  E  A  R  I  Q  V  W                   139
```

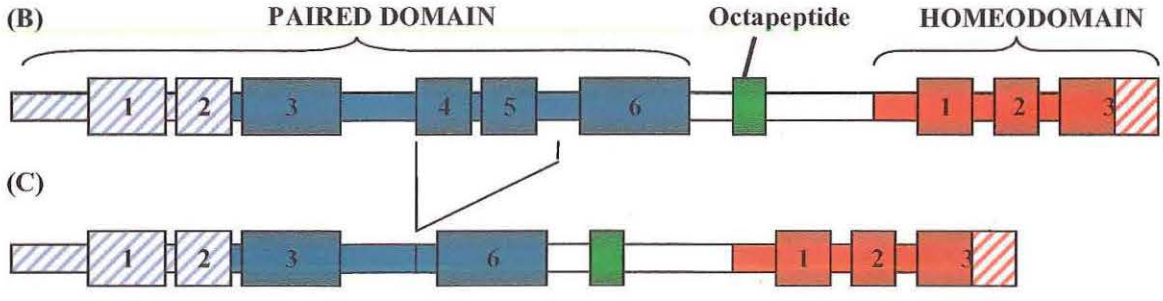


Figure 3.12: Molecular characterisation of alternate *Cdpax3/7-b* transcript/isoform. (A) Nucleotide and deduced amino acid sequence of partial (a) *Cdpax3/7-a* and (b) *Cdpax3/7-b* cDNAs; boxed sequence corresponds to the conserved paired box/paired domain (blue), homeobox/homeodomain (red) and octapeptide (green) sequence. Also shown is a schematic comparison of (B) Cdpax3/7-a and (C) Cdpax3/7-b protein structure showing α-helices of the PD and HD; hashing represents predicted regions (not sequenced); gaps in sections indicate lack of certainty about length.

4.0 DISCUSSION

Vertebrate *Pax3* and *Pax7* specify myoblast lineages (Seale *et al.*, 2000; Goulding *et al.*, 1994), and the emerging scenario suggests that *Pax7* may have a role in the conversion of certain stem cells into myoblasts (Polesskaya *et al.*, 2003; Seale *et al.*, 2000). These data, coupled with the high degree of evolutionary conservation of these genes (Noll, 1993), make *Pax* group III (*pgIII*) genes appropriate candidates for the study of the extensive adult muscle regeneration observed in decapod crustaceans such as *Cherax destructor*. As *Pax* genes had yet to be isolated within the Crustacea, the first part of this research concerned the isolation of *pgIII* genes from *C. destructor*. In itself this is important research, as the arthropod evolutionary lineage of *pgIII* genes is far from clear; extensive *pgIII* gene duplication has occurred within Arthropoda, however the positioning of these duplications is not yet known (Davis *et al.*, 2001). Thus knowledge of how many *pgIII* genes may have been present in early arthropods was not available above Chelicerata (Dearden *et al.*, 2002; Davis *et al.*, 2001).

The partially degenerate nested primers developed, and the methods utilised, in this research have facilitated the isolation of a *pgIII* gene from the Australian freshwater crayfish *C. destructor*. Comparison of the amino acid composition of the deduced protein sequence and phylogenetic analysis with other representative Pax proteins indicates that Cdpax3/7 is unequivocally a pgIII protein. Cdpax3/7 contains two putative DNA binding domains present in other pgIII proteins, the paired domain (PD) and the homeodomain (HD). The PD of Cdpax3/7 is highly similar to PDs of other pgIII proteins, showing 84% amino acid identity with Pby1 from the grasshopper *S. americana* and 76% amino acid identity with murine Pax3. The HD region shows similar homologies. As well as these conserved DNA binding regions, Cdpax3/7 also contains the octapeptide (HSIDGILG) motif, entirely conserved in Gsb, Pby1, Pax7, and Amphipax3/7 (Davis *et al.*, 2001; Holland *et al.*, 1999; Schafer, 1994; Baumgartner *et al.*, 1987). Cdpax3/7 forms a distinct clade with other arthropod pgIII proteins, and is most closely related to Pby1 from *S. americana* (81% amino acid identity).

Cdpax3/7 is expressed both embryonically and in adult muscle and premuscle tissues. Although *pgIII* genes have been isolated from many animals of different phyla, this

research marks the first report of the expression of a *pgIII* gene during myogenesis (either embryonic or adult) outside of the phylum Chordata. Vertebrate *Pax3* (Bober *et al.* 1994) and amphioxus *Amphipax3/7* (Holland *et al.*, 1999) are both expressed during embryonic myogenesis, and vertebrate *Pax7* is expressed during adult myogenesis (Seale *et al.*, 2000). *pgIII* genes isolated from mite (Dearden *et al.*, 2002), grasshopper (Davis *et al.*, 2001), fruit fly (Zhang *et al.*, 1994; Kilcherr *et al.*, 1986), and jellyfish (Miller *et al.*, 2000) have not been reported to be expressed during either embryonic or adult myogenesis. Moreover, this is the first documentation of the expression of a *pgIII* gene during adult regeneration outside of vertebrates.

4.1 *C. destructor* has a Single *pgIII* Gene

Only a single *pgIII* gene was isolated from *C. destructor*. Although the possibility of other *pgIII* genes in *C. destructor* cannot entirely be ruled out, several points are worthy of note. RT-PCR assays of 30% and 50% *C. destructor* embryos and adult muscle tissue did not display any evidence for the expression of *pgIII* genes other than *Cdpax3/7*. Thus if *C. destructor* is in possession of other *pgIII* genes, they would have to either not be expressed in 30% or 50% embryos and contain large intronic sequences (so as to not be amplified by genomic DNA PCR), or possess a sequence that was not amplified by the partially degenerate primers used for this research. As all other *pgIII* genes characterised in arthropod species to date are expressed in embryos at 30% and/or 50% of development (Dearden *et al.*, 2002; Davis *et al.*, 2001; Zhang *et al.*, 1994; Gutjahr *et al.*, 1993; Kilcherr *et al.*, 1986), it would be inconsistent for a *C. destructor pgIII* gene to not be expressed during these stages. The sites to which primers used in this research bind are regions that are highly conserved throughout arthropod *pgIII* sequences, and five different sets of primers were utilised (two of which were successfully used to isolate *pgIII* sequences from other arthropods, DF1/DR1 and DF2/DR2; Dearden *et al.*, 2002; Davis *et al.*, 2001). Thus it is unlikely that *pgIII* genes in *C. destructor* were not isolated due to the primers not amplifying them.

In addition, only a single *pgIII* gene has been found in Chelicerata (Dearden *et al.*, 2002). The maximum likelihood phylogenetic analysis presented here positions chelicerate *Tupax3/7* and crustacean *Cdpax3/7* in an ancestral location to all insect *pgIII* genes, thus no

support is given to the existence of other *pgIII* genes in the nearest common ancestor to these three classes. Therefore there is no evolutionary evidence for multiple *pgIII* genes being present in Crustacea. Due to the sum of this evidence, I am confident that *C. destructor* is in possession of only one *pgIII* gene.

4.1.1 *Cdpax3/7* Alternate Transcripts

Cdpax3/7 is expressed as two alternate transcripts. Alternate transcript *Cdpax3/7-b* encodes a protein with a 31 amino acid deletion. This deletion is within the RED domain of the PD of Cdpax3/7, and corresponds to the deletion of helices 5 and 6 of the PD. Alternatively spliced forms have not been detected for *D. melanogaster Prd, Gsb*, or *Gsb-n*, *T. urticae Tupax3/7*, *A. millepora PaxD*, or *B. floridae Amphipax3/7* (Dearden *et al.*, 2002; Davis *et al.*, 2001; Miller *et al.*, 2000; Holland *et al.*, 1999). Vertebrate *Pax3* and *Pax7* do generate alternatively spliced forms, resulting in the inclusion of one, three, or five amino acids into the PD (Seo *et al.*, 1998; Ziman & Kay, 1998). *S. americana Pby1* and *Pby2* genes are the only other invertebrate *pgIII* genes, thus far identified, to generate alternatively spliced forms (Davis *et al.*, 2001). Alternate splicing results in the inclusion of five amino acids in the PD of Pby1, and the deletion of three amino acids from the HD of Pby2 (Davis *et al.*, 2001). The size of the alternate transcript deletion (31 amino acids) identified in the PD of *Cdpax3/7* from *C. destructor* is far greater than those identified in any of its homologues.

The genomic structure of *Cdpax3/7* was not investigated in this research; however, from the data presented here it is likely that the *Cdpax3/7* gene contains at least two introns located between the 5′ end of the paired box and the 3′ end of the homeobox. The primers flanking this region used in this study did not amplify *Cdpax3/7* from *C. destructor* genomic DNA, indicating that the presence of introns resulted in the amplification region being too large for this technique (approximately 4kb; Qiagen, 2003). The presence of two alternate transcripts of *Cdpax3/7* is also indicative of the presence of introns. The size of the deletion present in *Cdpax3/7-b*, a 93 nucleotide region spliced from the paired box, indicates that this transcript is likely produced via exon-skipping (Miriami *et al.*, 2003). If this is the case, it would indicate that two introns are present within the paired box region of

Cdpax3/7, and that these introns are located approximately at the splice sites of *Cdpax3/7-a* and *Cdpax3/7-b* (corresponding to amino acids 75 and 105 of the PD).

4.1.2 *Cdpax3/7* Splice Site Locations

The splice site at amino acid position 105 of the PD of Cdpax3/7 is conserved in both *D. melanogaster* Gsb-n and *S. americana* Pby1 (Figure 4.1). Both *Gsb-n* and *Pby1* contain splice sites at positions corresponding to amino acid position 105 of the PD (Davis *et al.*, 2001; Gutjahr *et al.*, 1993). The location of intron 2 (4140bp) of the *Gsb-n* gene corresponds to amino acid position 105 of the PD (Gutjahr *et al.*, 1993). In this research, the *Gsb-n* gene did not result in amplification products by PCR from *D. melanogaster* genomic DNA, indicating that the inclusion of intron sequence resulted in the genomic *Gsb-n* amplification region exceeding the requirements of the technique. Thus it is likely that the inclusion of a similar intron sequence resulted in the lack of *Cdpax3/7* amplification from *C. destructor* genomic DNA. This splice site is not found in other *pgIII* genes (Dearden *et al.*, 2002; Davis *et al.*, 2001; Miller *et al.*, 2000; Holland *et al.*, 1999; Gutjahr *et al.*, 1993; Frigerio *et al.*, 1986).

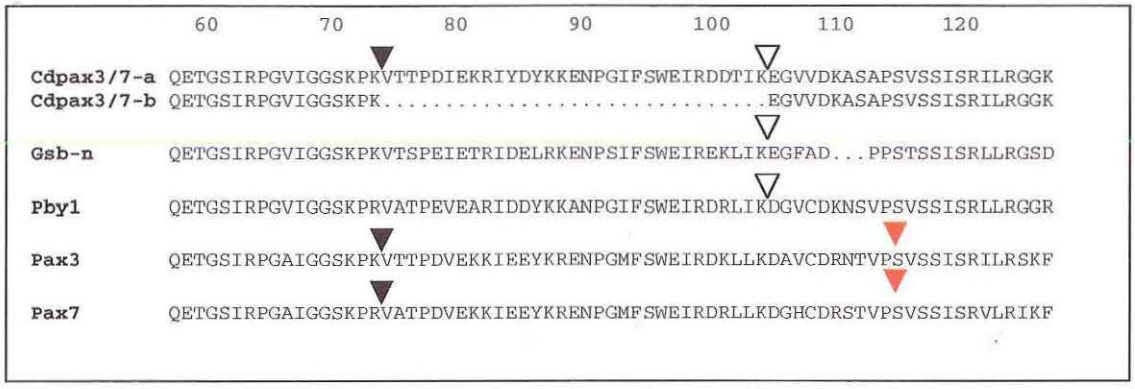


Figure 4.1: Location of splice sites in the PD of Cdpax3/7 and other Pax proteins. Black arrows indicate splice sites conserved between Cdpax3/7 and vertebrate Pax3 and Pax7, white arrows indicate splice sites conserved between Cdpax3/7 and arthropod Gsb-n and Pby1, red arrows indicate splice sites present only in vertebrate Pax3 and Pax7. Numbers above sequences indicate PD amino acid positions and dots (.) indicate amino acid deletions.

Both vertebrate *pgIII* genes *Pax3* and *Pax7* contain introns at the position corresponding to amino acid position 75 of the PD (Seo *et al.*, 1998; Ziman & Kay, 1998). Thus the amino acid position 75 splice site is conserved in *Cdpax3/7* and vertebrate *Pax3* and *Pax7* (Figure 4.1) (Ziman & Kay, 1998; Vogan *et al.*, 1996; Seo *et al.*, 1998). As both of the *Cdpax3/7* splice sites coincide with the presence of introns in *pgIII* genes of other animals, further confirmation is given to the idea that the *Cdpax3/7-b* alternate transcript is generated by exon-skipping. Exon skipping is a common method for alternate transcript generation in *M. musculus*, where skipped exons range in size from 12 to 236bp (Miriami *et al.*, 2003), a range which certainly encompasses the 93bp section spliced from *Cdpax3/7-b*.

The conservation of the position 75 splice site between *C. destructor* Cdpax3/7 and vertebrate Pax3/Pax7 is of particular interest. This intron/splice site at position 75 of the PD is not found in any of the invertebrate *pgIII* genes/cDNAs isolated to date (Dearden *et al.*, 2002; Davis *et al.*, 2001; Miller *et al.*, 2000; Holland *et al.*, 1999; Gutjahr *et al.*, 1993; Frigerio *et al.*, 1986). Interestingly, this intron/splice site is not present in the nearest common ancestral gene to both *Pax3* and *Pax7*, amphioxus *Amphipax3/7* (Holland *et al.*, 1999). *Amphipax3/7* does contain many intron locations that are conserved in vertebrate *Pax3* and *Pax7*, but *Amphipax3/7* does not contain an intron at position 75 of the PD (Holland *et al.*, 1999).

The conservation of this splice site between *Cdpax3/7* and vertebrate *Pax3* and *Pax7* presents an interesting evolutionary scenario. It indicates that either the ancestral *pgIII* gene to both crustaceans and vertebrates (prior to the protostome-deuterostome split) also contained this splice site, and that it has subsequently been lost in all other *pgIII* genes of all other species identified thus far, or the splice site has evolved twice in the evolutionary history of *pgIII* genes.

4.2 Functions and Expression Patterns of Arthropod *pgIII* Genes

D. melanogaster pgIII genes are involved in early body patterning and specification of the developing central nervous system (CNS) (Zhang *et al.*, 1994). *D. melanogaster pgIII* genes *Paired* (*Prd*) and *Gooseberry* (*Gsb*) were first classified as segmentation genes (Kilcherr *et al.*, 1986). *Prd* functions as a secondary pair-rule gene that specifies the

parasegmental organisation of the developing embryonic cuticle (Bertuccioli *et al.*, 1996; Kilcherr *et al.*, 1986; reviewed in Davis & Patel, 1999) and is also necessary for male fertility and survival to adulthood (Xue & Noll, 2002; Xue *et al.*, 2001). *Gsb* is a segment polarity gene (Li & Noll, 1994). *Gsb* is activated by pair-rule genes (specifically *Prd*) and its expression defines pattern within each cuticle segment (Bouchard *et al.*, 2000; Duman-Scheel *et al.*, 1997; Gutjahr *et al.*, 1993).

Both *Gsb* and the third *D. melanogaster pgIII* gene *Goosberry-neuro* (*Gsb-n*) are also vital for CNS development (Li & Noll, 1994; Baumgartner *et al.*, 1987). *Gsb* and *Gsb-n* specify the posterior regions of developing rows of neuroblasts in the CNS (Bouchard *et al.*, 2000; Duman-Scheel *et al.*, 1997; Zhang *et al.*, 1994). In the developing *D. melanogaster* embryo, *Gsb* expression is initially restricted to the ectoderm and neuroectoderm, but is later expressed in mesodermal and neuronal cells (Zhang *et al.*, 1994; Gutjahr *et al.*, 1993; Baumgartner *et al.*, 1987). Whilst the function of the *Prd* gene has been relatively exhaustively studied, many of the later functions of *Gsb*, and particularly *Gsb-n*, remain to be elucidated (Bouchard *et al.*, 2000; Duman-Scheel *et al.*, 1997; Gutjahr *et al.*, 1993).

The precise functions of other invertebrate *pgIII* genes (*Pby1*, *Pby2*, *Tupax3/7*, *PaxD*) have not been studied (Dearden *et al.*, 2002; Davis *et al.*, 2001; Miller *et al.*, 2000). In the developing embryo of the grasshopper *S. americana*, *pgIII* genes *Pby1* and *Pby2* are expressed in patterns similar to the combined expression of the three *D. melanogaster pgIII* genes; as stripes in the segmenting epidermis and in the developing CNS (Davis *et al.*, 2001). Mite *Tupax3/7* is also expressed as stripes in the early developing epidermis (during segmentation) and in the developing CNS (Dearden *et al.*, 2002).

It is of interest to this research that both *Pby1* and *Pby2* are expressed in rings in developing limb primordia in *S. americana* (Davis *et al.*, 2001) and that *Tupax3/7* is also expressed in rings in the developing limbs of *T. urticae* embryos, and of newly hatched nymphs (Dearden *et al.*, 2002). These expression patterns are not appropriately accounted for by the known expression patterns of *D. melanogaster pgIII* genes (Bouchard *et al.*, 2000; Duman-Scheel *et al.*, 1997; Gutjahr *et al.*, 1993; Kilcherr *et al.*, 1986; Baumgartner *et al.*, 1987); however this expression may be of interest in the patterning of limb muscle.

4.3 The Expression of *Cdpax3/7* during Adult Myogenesis in *C. destructor*

Cdpax3/7 is expressed in the adult claw closer muscle tissue of *C. destructor* throughout the entire moult cycle (intermoult, premoult and postmoult). The claw closer muscle of *C. destructor*, and certain other decapod crustaceans, undergoes substantial atrophy and regeneration during stages of the moult cycle (Lamey *et al.*, 2002; Govind & Pearce, 1994; Mykles & Skinner, 1982). In premoult, the claw closer muscle of *C. destructor* atrophies by as much as 60%, and following ecdysis rapidly regenerates and grows (Mykles & Skinner, 1982), whereas intermoult is a period of myogenic quiescence (West *et al.*, 1995). Cellular studies in the snapping shrimp *Alpheus heterochelis* indicate that muscle satellite cells act as myogenic precursors for muscle regeneration in the moult cycle regeneration process (Govind & Pearce, 1994). The activation of satellite cells in this process is believed to occur during the premoult phase and the satellite cells affect regeneration throughout the postmoult regenerative phase (Holland & Skinner, 1976). So *Cdpax3/7* is expressed in claw closer muscle tissue during myogenic quiescence and during periods of satellite cell activation and regeneration.

Cdpax3/7 is also expressed during epimorphic regeneration of the cheliped, during both papilla and closed dactyl stages. Studies in other decapod crustaceans (*A. heterochelis, Astacus fluviatilis, Uca pugilator*) indicate that in this process a blastema is first formed by a proliferation of epithelial cells, fibroblasts then provide a structural framework, followed by a huge influx of haematopoietic stem cells (Hopkins *et al.*, 1999; Read & Govind, 1998; Erri Babu, 1987). The immigrant haematopoietic stem cells provide an important source of myoblasts in the myogenesis that follows (Read & Govind, 1998; Pearce *et al.*, 1997; Uhrik *et al.*, 1989; Erri-Babu, 1987). The phenomenal capability for almost limitless adult myogenesis in *C. destructor*, and other decapod crustaceans, appears to result from their increased ability to utilise haematopoietic stem cells as myoblasts (Read & Govind, 1998; Pearce *et al.*, 1997; Uhrik *et al.*, 1989; Erri Babu, 1987).

In *C. destructor*, epimorphic regeneration appears to be completed towards the end of the closed dactyl stage (Cutler *et al.*, 2002); indicating that the myogenic determination of immigrant haematopoietic stem cells is occurring during the papilla stage (Cutler *et al.*, 2002). Taken with the expression patterns, this suggests a role for *Cdpax3/7* in normal

non-regenerating (intermoult) adult muscle, muscle undergoing satellite cell-directed regeneration (premoult and postmoult muscle), as well as during the conversion of non-satellite stem cells into muscle (the conversion of haemocytes into muscle) that occurs during epimorphic regeneration.

The expression of *Cdpax3/7* in each of the corresponding cell lineages was beyond the scope of this project, but would likely lead to significant findings. However, certain inferences about the cellular environment can be deduced from known differences in the cellular composition of the tissues in which *Cdpax3/7* is expressed. The intermoult, premoult and postmoult samples consist of myofibres along with their associated satellite cells. Thus we may hypothesise that *Cdpax3/7* is expressed in either myofibres or satellite cells, or possibly both, in the claw closer muscle of *C. destructor*.

The cellular environment of epimorphic regeneration is quite different from that of moult-cycle regenerating muscle and several types of cells are present, including epithelial cells, fibroblasts, and haemocytes (Read & Govind, 1998). The cell types also differ between the papilla and closed dactyl stages, in that by the closed dactyl stage myofibres (and presumably satellite cells) are present (Cutler *et al.*, 2002). The significant point, however, is that no muscle fibres are present in epimorphically regenerating limb at the papilla stage, but they are recognised at the closed dactyl stage (Cutler *et al.*, 2002). *Cdpax3/7* is expressed at both of these stages of epimorphic regeneration, indicating that *Cdpax3/7* is expressed in undifferentiated tissue undergoing myogenesis prior to the formation of myofibres. These gross expression patterns may signify that *Cdpax3/7* is expressed in quiescent (intermoult) and activated (premoult/postmoult) satellite cells, and in stem cells undergoing myogenic determination (haemocytes in the epimorphically regenerating papilla); however cell specific studies are required to confirm this.

4.3.1 Comparison of the Myogenic Expression of *Cdpax3/7* to Vertebrate *Pax7*

The cellular processes involved in moult cycle muscle regeneration in *C. destructor* and adult muscle regeneration in higher vertebrates show a great deal of similarity. Crustacean satellite cells are analogous to those of vertebrates (Uhrik *et al.*, 1989; Novotova & Uhrik, 1992; Pearce *et al.*, 1997). In response to signals, quiescent satellite cells are activated,

migrate and fuse to a myofibre (or each other) effecting regeneration. This process is highly regulated in both higher vertebrates and in *C. destructor*. Muscle side population (SP) cells are known to be myogenic progenitors in adult muscle of mice (Polesskaya *et al.*, 2003; Asakura *et al.*, 2002). Muscle SP cells have not yet been identified in crustacean muscle, however haematopoietic stem cells may be utilised in muscle regeneration in both vertebrates (LaBarge & Blau, 2002) and crustaceans (Uhrik *et al.*, 1989).

The amount of haematopoietic stem cells actually utilised during normal muscle regeneration in adult vertebrates, however, is predicted to be very small (reviewed in Grounds *et al.*, 2002). In contrast, crustaceans appear to utilise haematopoietic stem cells freely during moult cycle muscle regeneration (Uhrik *et al.*, 1989), and specifically during epimorphic regeneration, where haematopoietic stem cells appear to be the dominant myogenic precursor cell (Read & Govind, 1998; Pearce *et al.*, 1997; Uhrik *et al.*, 1989; Erri Babu, 1987).

In mice, *Pax7* has been found to specify satellite cells (Seale *et al.*, 2000) and also to be involved in the conversion of muscle SP cells into satellite cells (Asakura *et al.*, 2002) in response to Wnt signals (Polesskaya *et al.*, 2003). Haematopoietic stem cells injected into mice have been shown to occupy the satellite cell compartment (LaBarge & Blau; 2002), a cell type that is specified by *Pax7* (Seale *et al.*, 2000).

The gross expression pattern of *Cdpax3/7* in adult muscle of *C. destructor* displays similarities to the gross expression patterns of *Pax7* in adult muscle of mice/higher vertebrates. In mice, *Pax7* is expressed in normal (non-regenerating) adult muscle (specifically in quiescent satellite cells; Seale *et al.*, 2000), muscle undergoing satellite cell-driven regeneration (specifically in activated satellite cells; Seale *et al.*, 2000), as well as during the conversion of non-satellite stem cells into muscle (the conversion of muscle side population [SP] cells into satellite cells; Polasskaya *et al*, 2003). These gross expression patterns are somewhat analogous to that of *Cdpax3/7* in intermoult, premoult, and postmoult claw closer muscle and regenerating cheliped at the papilla and closed dactyl stages.

The expression of *Cdpax3/7* in *C. destructor* epimorphically regenerating cellular tissue in which haematopoietic stem cells are undergoing myogenic determination (Read & Govind, 1998; Pearce *et al.*, 1997; Uhrik *et al.*, 1989; Erri Babu, 1987) is of great interest to the study of *pgIII* gene function. Whilst it is not possible to draw conclusions about the specific cellular expression or function of *Cdpax3/7* from this research, it is tempting to speculate that *Cdpax3/7* may specify muscle satellite cells, and possibly be involved in the conversion of non-satellite stem cells to the myogenic lineage. The expression patterns of *Cdpax3/7* presented here are certainly not in conflict with this notion.

4.4 Cdpax3/7 Protein Structure and DNA Binding Function

As transcription factors, the function of Pax proteins depends upon their ability to recognise and bind specific DNA sequences in the promoter/enhancer regions of downstream tissue-specific target genes. This ability to recognise DNA sequences is a direct consequence of the structure of the pgIII protein, therefore to deduce putative functions of Cdpax3/7, it is first necessary to compare the structure of Cdpax3/7 to other pgIII proteins of known function.

4.4.1 DNA Binding Properties

Pax proteins are capable of utilising either or both PD and HD DNA binding domains in target DNA sequence recognition (Xu *et al.*, 1999; Jun *et al.*, 1998; Miskiewicz *et al.*, 1996; Jun & Desplan, 1996; Xu *et al.*, 1995). Also, due to the bipartite nature of the PD, either or both of its domains (PAI and RED) are capable of being used in DNA binding (Xu *et al.*, 1999; Jun *et al.*, 1998; Kozmik *et al.*, 1997; Jun & Desplan, 1996; Xu *et al.*, 1995). This inherent modularity gives certain Pax proteins the ability to recognise disparate DNA sequences and also bind DNA with differing affinity (Jun & Desplan, 1996). The amount to which these three binding domains are utilised depends on the positioning of conserved DNA binding and structural amino acids derived from the particular gene (Apuzzo & Gros, 2002; Sun *et al.*, 2002), but may be altered by post-transcriptional modifications, the most important of which is alternate splicing of mRNA (Kozmik *et al.*, 1997; Tell *et al.*, 1998). Alternate splicing leads to proteins with insertions or deletions that may alter the conformation or chemical properties of the final transcription factor (Miriami *et al.*, 2003).

4.4.2 PAI Domain of the PD

The PAI domain is the dominant region responsible for DNA binding affinity and specificity of the majority of Pax proteins (Kozmik *et al.*, 1997). Helix 3 of the PAI domain fits directly within the DNA major groove and participates in DNA sequence recognition by amino acid contact of specific DNA bases (Figure 4.2) (Xu *et al.*, 1999; Xu *et al.*, 1995), thus helix 3 is designated the recognition helix of the PAI domain. Helix 3 of the RED domain (Figure 4.2) and helix 3 of the HD both fit directly within the DNA major groove and participate in DNA sequence recognition by amino acid contact of specific DNA bases (Xu *et al.*, 1999; Wilson *et al.*, 1995) and are both designated as the recognition helices of their specific DNA binding domains.

The amino acid composition of the paired domains of Pax proteins is highly conserved, yet different Pax proteins function distinctly by binding specifically to different recognition sequences (Jun & Desplan, 1996). Amino acids at certain positions exert a disproportionate effect upon the binding specificities of particular proteins (Apuzzo & Gros, 2002). Prd and PAX6 proteins have been shown to bind *in vitro* to different recognition sequences (Sun *et al.*, 2002) and have different *in vivo* functions (Sun *et al.*, 2002; Gutjahr *et al.*, 1993). Position 47 of the PD is the first amino acid of the PAI domain recognition helix (helix 3), and has been found to contact the DNA recognition sequence at the major groove in both Prd and PAX6 crystal structures (Xu *et al.*, 1999; Xu *et al.*, 1995). PAX6 contains an asparagine residue at position 47 whereas Prd contains a histidine residue at position 47. These residues have been shown to interact differently with DNA, accounting for the recognition of different DNA sequences by the two proteins (Jun & Desplan, 1996). His-47 of Prd hydrogen bonds to a guanine in its recognition sequence (Xu *et al.*, 1995), whereas Asp-47 of PAX6 makes van der Waals contact with a thymine (Xu *et al.*, 1999).

Further evidence for the pivotal nature of amino acids at position 47 in the binding specificity of Pax proteins has come from reciprocal mutations of Prd and PAX6 proteins. Jun and Desplan (1996) demonstrated that mutating Prd to carry Asp-47 (Prd-Asp-47) induced a 23-fold increase in binding to a PAX6 recognition sequence, and *vice versa* (with a 7-fold increase) for PAX6-His-47. PAX6 and its evolutionary homologues are master regulatory genes in eye development in many animals, including both vertebrates and *D.*

melanogaster (Tell *et al.*, 1998). An ectopic eye assay has been used to assess the functional significance of having asparagine or histidine at position 47. The ectopic expression of an *eyeless* (*D. melanogaster PAX6* homologue containing Asp-47) transgene results in the formation of supernumerary (ectopic) eyes (Halder *et al.*, 1995). Mutating *eyeless* to carry His-47 results in a 75% reduction in the ectopic eye phenotype (Sun *et al.*, 2002), indicating that position 47 has a dominant effect in binding site recognition.

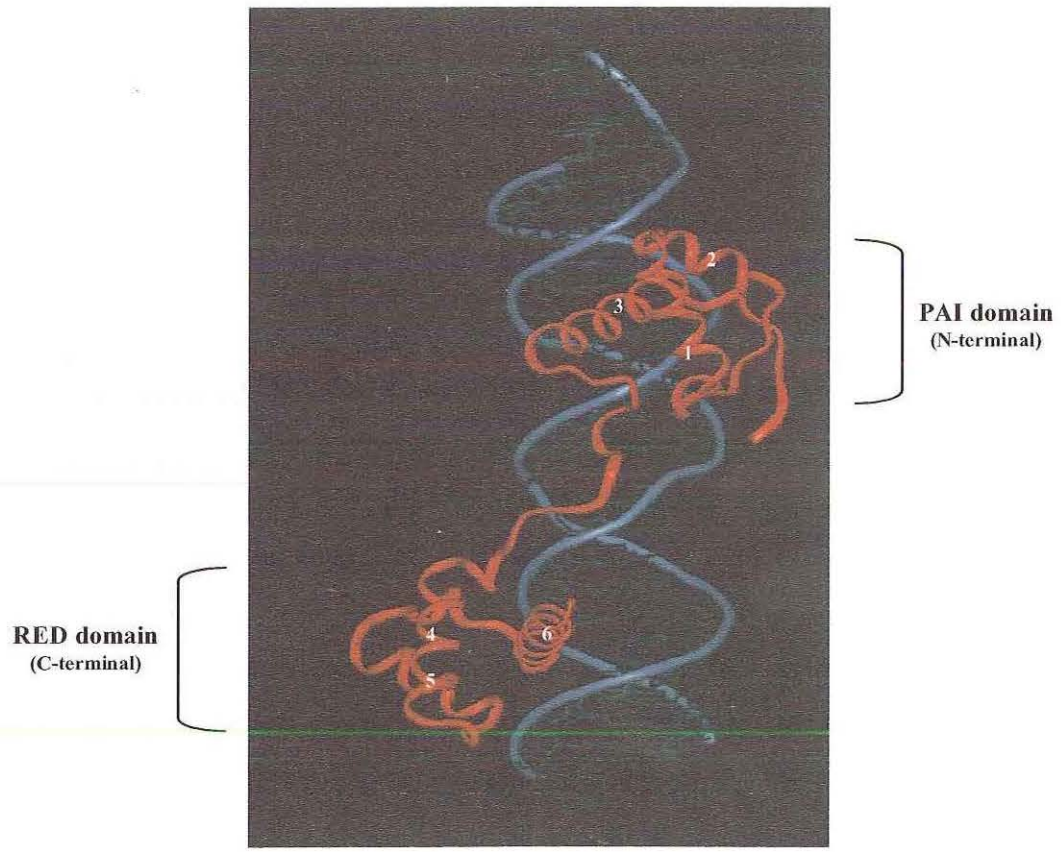


Figure 4.2: Structure of *Homo sapiens* PAX6 protein (red) binding to DNA (blue) showing conserved PAI (α-helices 1 - 3) and RED (α-helices 4 - 6) DNA binding domains (Xu *et al.*, 1999). Of note are the recognition helix of the PAI domain (helix 3) and the recognition helix of the RED domain (helix 6) and their locations within the DNA major grooves.

Cdpax3/7 exhibits 100% homology of amino acids at positions that contact DNA in the Prd crystal structure, and Cdpax3/7 carries a histidine residue at position 47 of its PD. This provides strong evidence for the Cdpax3/7 PAI domain binding similar recognition sequences to the Prd PAI domain. This indicates that Cdpax3/7 may be capable of

sequences to the Prd PAI domain. This indicates that Cdpax3/7 may be capable of activating similar downstream target genes to Prd, and through this may have similar functions.

4.4.3 RED Domain of the PD

The RED domain of Prd protein has not been found to make contact with its preferred DNA recognition sequence and is dispensable for normal function (Xu *et al.*, 1995; Bertuccioli *et al.*, 1996). A major difference between PAX6 (which utilises its RED domain) and Prd (which does not utilise its RED domain) is amino acid residue 119. PAX6 contains a serine (polar) residue at position 119, whereas Prd has an alanine residue (non-polar) at this position. The crystal structure of PAX6 shows that Ser-119 hydrogen bonds with oxygen present in the sugar-phosphate backbone of its DNA recognition sequence (Xu *et al.*, 1999). The Ala-119 present in Prd is hydrophobic, and therefore, not only does it not create this important hydrogen bond to the sugar-phosphate backbone, it repels the DNA by placing a hydrophobic group near a phosphate. This major difference is postulated to be the prime causal factor in Prd not utilising its RED domain in DNA recognition and binding (Xu *et al.*, 1999). Cdpax3/7 contains a serine residue at position 119 of its PD, and thus does not have the same hydrophobic interaction as Prd. Murine Pax3 and Pax7 also possess the Ser-119 residue, and both of these proteins utilise their RED domains to some degree in DNA recognition and binding (Vogan *et al.*, 1996).

An alternate splicing event resulting in the insertion of a glutamine residue at the N-terminal end of the RED domain (Q+) only mildly modifies the DNA-binding properties of Pax3 and Pax7, whereas the equivalent insertion in PAX6 drastically affects protein function (Kozmik *et al.*, 1997; Vogan *et al.*, 1996). The additional glutamine residue weakens the DNA contacts made by the RED domain in all three proteins (Vogan *et al.*, 1996), so the effects on function can be taken as indicating that PAX6 is more reliant upon its RED domain than either Pax3 or Pax7, which are only mildly reliant on RED domain DNA binding for function (Kozmik *et al.*, 1997). This is different from the case of Prd, which does not bind at all with its RED domain (Xu *et al.*, 1995). *In vivo* studies of altered RED domains show normal function in Prd, but show altered function in Pax3 and Pax7, indicating that these proteins normally utilise their RED domains to a varied degree

(Bertuccioli *et al.*, 1996; Vogan *et al.*, 1996). As the main difference between these proteins' PDs may be put down to Ser-119, and Cdpax3/7 shares Ser-119 with Pax3 and Pax7, it is likely that Cdpax3/7 is at least capable of utilising its RED domain for DNA binding, possibly in a similar manner to the two vertebrate pgIII proteins. This may indicate that the entire PD sequence (PAI and RED) of Cdpax3/7 is capable of binding similar DNA sequences to those bound by vertebrate Pax3 and Pax7. Thus Cdpax3/7 may be capable of binding and activating similar downstream target genes to Pax3 and Pax7, and through this may have similar function.

4.4.4 The Significance of the Alternate Isoform Cdpax3/7-b

The alternate splicing event leading to the alternate isoform Cdpax3/7-b eliminates the first two α-helices of the RED domain, from amino acid position 75 to 105 of the PD. This results in a protein with a shortened PD with the PAI domain recognition helix (α-helix 3) and the RED domain recognition helix (α-helix 6) in close proximity. It is unlikely that the spatial orientation provided by this shortened isoform would allow the RED domain recognition helix (α-helix 6 of the PD) to reach the major groove of DNA in the same manner that it does in the crystal structure of PAX6 (Figure 4.2; Xu *et al.*, 1999). Therefore this alternate transcript may result in a protein that binds in a similar manner to Prd, which does not utilise its RED domain for DNA sequence recognition (Xu *et al.*, 1995). Elimination of two thirds of the RED domain could completely ablate any DNA-binding function that the RED domain may have.

If the consequence of this alternate splicing event is the production of a shortened protein (Cdpax3/7-b isoform) that utilises the same binding domains as Prd, and a normal protein (Cdpax3/7) that binds utilising the same binding domains as Pax3 and Pax7 (and presumably Gsb and Gsb-n, both of which also contain the Ser-119 residue), then this may allow *Cdpax3/7* to perform all of the DNA-binding functions consistent with an ancestral gene to all of its descendant paralogues. Interestingly, the two pgIII proteins isolated from the grasshopper *S. americana*, Pby1 and Pby2, differ at amino acid 119. Pby2 has an Asp-119 residue whereas Pby1 has a Ser-119 residue (Davis *et al.*, 2001). If indeed Asp-119 does preclude the use of the RED domain in DNA binding, Pby2 may bind to recognition sequences characteristic of Prd, whereas Pby1 may bind to recognition sequences

characteristic of Gsb/Gsb-n. Having pgIII proteins that utilise their entire PD for binding as well as pgIII proteins that only use their PAI domain for binding may be an important (even conserved) occurrence for the maintenance of functions performed by *pgIII* genes in the arthropod lineage. An alternate splicing event such as that observed in Cdpax3/7 may then allow a single gene to perform all of these functions.

4.4.5 Alternate Splicing within the RED Domain Alters DNA Binding

Pax3 and Pax7 proteins are capable of recognising DNA sequences by the use of either their PAI, RED, or HD domains, or any combination of the three (Apuzzo & Gros, 2002; Jun & Desplan, 1996). DNA binding assays have identified DNA sequences that are bound by each of the three domains of Pax proteins. P2 sequences are recognised entirely through HD interaction (Underhill & Gros, 1997), e5 sequences are recognised entirely through PAI domain interaction (Czerny *et al.*, 1993; Xu *et al.*, 1995; Vogan *et al.*, 1996), 5aCON sequences are only recognised through the RED domain (Kozmik *et al.*, 1997), and P6CON and CD19/2 sequences are only bound by interaction of both PAI and RED domains (Chalepakis & Gruss, 1995; Vogan *et al.*, 1996). The preferential binding of these different recognition sequences is heavily regulated by protein isoforms generated by alternate splicing.

The ability of Pax3 and Pax7 proteins to utilise their RED domains is modulated by an alternate splicing event in the linker region between the PAI and RED domains of the PD (Underhill & Gros, 1997; Vogan & Gros, 1997). The inclusion of an additional glutamine residue at position 75 (designated Q+) of the PD reduces the binding affinity of the RED domain between two- and five-fold, and alters the recognition sequence specificity of the resultant protein (Vogan *et al.*, 1996). Q- isoforms bind sequences utilising PAI, RED and HD domains (P6CON, CD19/2, and e5 recognition sequences) whereas Q+ isoforms only utilise PAI and HD domains in DNA binding (binding only to P2 and e5 recognition sequences) (Vogan *et al.*, 1996). Both isoforms are present in *Pax3* and *Pax7* expressing tissues of higher vertebrates, and interestingly it is this Q+ isoform that is most abundant (Ziman & Kay, 1998; Vogan *et al.*, 1996; Seo *et al.*, 1998).

The other alternate splicing event of *Pax3* and *Pax7* results in the inclusion of glycine (G) and leucine (L) residues at position 118 (between helix-5 and helix-6) of the PD (designated GL+). Changes in secondary structure of the RED domain recognition helix (helix 6) in the GL+ isoform are predicted to alter binding specificity (Ziman & Kay, 1998). The GL+ isoform is predicted to reduce RED domain binding in the same way as Q+ isoform (Ziman & Kay, 1998), so the Q+/GL+ isoform is likely to be unable to bind with its RED domain.

The alternate transcript identified in *C. destructor Cdpax3/7-b* produces a protein with a rather more severe alteration than all of those identified above for its homologues, Pax3 and Pax7. Cdpax3/7-b lacks two-thirds of its RED domain, indicating that it has lost important DNA binding amino acids within this region and that its tertiary structure is also severely altered. Alteration of pgIII protein DNA binding function through alternate splicing alterations in the RED domain of the PD appears to be an important mechanism for regulating the different DNA binding functions of Pax3 and Pax7 (Vogan *et al.*, 1996; Ziman & Kay, 1998). RED domain alternate splicing may, therefore, also be important for regulation of the functions of Cdpax3/7 in *C. destructor*.

4.5 The Evolution of *pgIII* Genes, Transcriptional Hierarchies and Myogenesis

4.5.1 The Evolution of *Pax* Group III Genes

This research marks the first description of a *pgIII* gene in Crustacea, and in doing so adds to our understanding of the protostome evolutionary lineage of *pgIII* genes. Phylogenetic studies place the Chelicerata as the basal arthropod class, with Hexapoda and Crustacea forming a latter-diverging sister group (Figure 4.3; Turbeville *et al.*, 1991; reviewed in Davis & Patel, 1999). The description of *Cdpax3/7* gives us greater certainty in the positioning of the duplications giving rise to the *pgIII* genes we find in extant insect species. Phylogenetic analysis places *Cdpax3/7* as a direct orthologue (separated by speciation, not duplication) to the ancestral *pgIII* gene. This gives a single *pgIII* genetic lineage from the Cnidaria-higher Metazoa split, through the protostome-deuterostome split and the divergence of the Chelicerata from other Arthropoda, to the time of the divergence of Crustacea from Hexapoda (Figure 4.4).

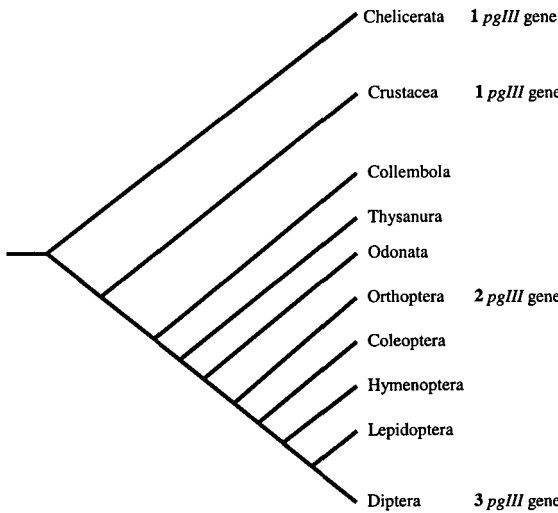


Figure 4.3: Arthropod evolutionary tree showing known *pgIII* genes (evolutionary tree adapted from Davis & Patel, 1999; gene positioning from Dearden *et al.*, 2002; Davis *et al.*, 2001; Baumgartner *et al.*, 1987; Kilcherr *et al.*, 1986; and this study).

Extensive *pgIII* proliferation has occurred within arthropods. Maximum likelihood analysis (this paper) and both maximum parsimony and maximum likelihood analysis (Dearden *et al.*, 2002; Davis *et al.*, 2001) unequivocally indicate that substantial *pgIII* gene duplication has occurred within the Arthropoda. The occurrence of only a single *pgIII* gene in the chelicerate *T. urticae* and in the crustacean *C. destructor*, coupled with the maximum likelihood analysis presented here, presents a strong case for the existence of only a single *pgIII* gene prior to the divergence of the Crustacea and the ancestor of Insecta (greater than 511 Million years ago; Siveter *et al.*, 2001).

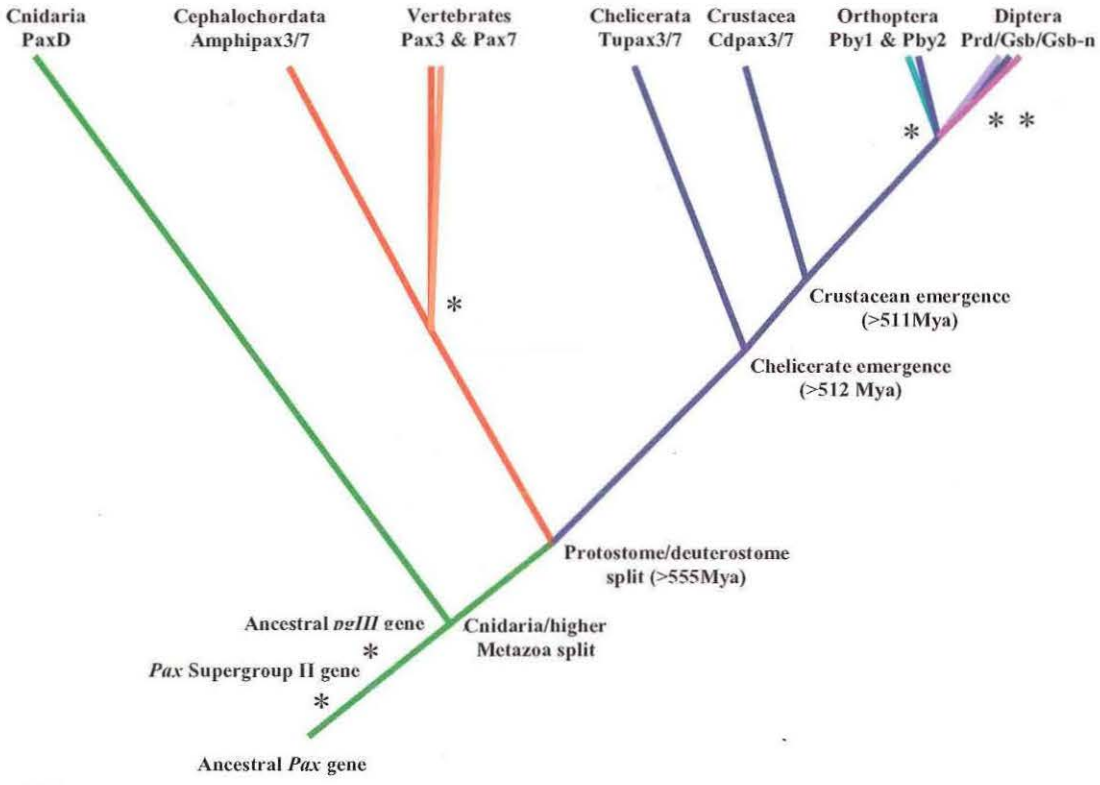


Figure 4.4: Evolutionary tree showing positions of duplications of *Pax* group III genes (and progenitor *Pax* genes). Positions of duplications are indicated by asterisks, multiple lines indicate multiple *pgIII* paralogous genes. Green lineage indicates genes present in pre-bilaterian fauna, red lineage indicates genes present in deuterostome fauna, blue lineage indicates genes present in protostome fauna. Tree topology is that of animal evolution outlined in Holland (2003), Davis and Patel (2002), Turbeville *et al.* (1991) and Briggs and Collins (1988). Early *Pax* duplications are those inferred by Sun *et al.* (1997) and Miller *et al.* (2001). Positioning of genes shown are those indicated by Miller *et al.* (2001), Holland *et al.* (1999), Dearden *et al.* (2002), Davis *et al.* (2001), and this study. It should be noted that the exact positioning of the duplications occurring within the Insecta have not been resolved (Davis *et al.*, 2001; this study).

The description of *Cdpax3/7* thus defines the placement of the arthropod *pgIII* duplications as occurring after the divergence of Crustacea from Hexapoda. While this paper and others (Dearden *et al.*, 2002; Davis *et al.*, 2001) have not completely resolved the phylogenetic radiation of pgIII proteins within the Insecta (due to insignificant bootstrap support values for branches between *S. americana* Pby1/Pby2 and *D. melanogaster* Prd/Gsb/Gsb-n in all cases; Dearden *et al.*, 2002; Davis *et al.*, 2001), there is some support for this pgIII proliferation to have occurred after the divergence of insect orders Orthoptera (*S. americana*) and Diptera (*D. melanogaster*), as Pby1 and Pby2 share higher identity to each

cases; Dearden *et al.*, 2002; Davis *et al.*, 2001), there is some support for this pgIII proliferation to have occurred after the divergence of insect orders Orthoptera (*S. americana*) and Diptera (*D. melanogaster*), as Pby1 and Pby2 share higher identity to each other than to any of the *D. melanogaster* pgIII proteins (Davis *et al.*, 2001). The occurrence of three *pgIII* genes in *D. melanogaster* and two *pgIII* genes in *S. americana* is thus attributed to significant *pgIII* radiation within the Insecta, whether this had occurred prior to the *D. melanogaster*/*S. americana* split is not resolved.

4.5.2 The Nearest Common Ancestor to *C. destructor* and *M. musculus*

Evidence of the last common protostome-deuterostome ancestor has yet to be found in the fossil record. Metazoa first appear in the fossil record in the Ediacaran and Duoshanto deposits, originating some 570 Million years ago (Mya) (Erwin & Davidson, 2002). Cnidaria and Porifera are both represented (Li *et al.*, 1998; Chen *et al.*, 2002). The first bilaterian (*Kimberella*) appeared 555 Mya (Martin *et al.*, 2000). However, *Kimberella* exhibits a high affinity with protostomes (it is mollusc-like), leading many researchers to predict that the protostome-deuterostome divergence had occurred prior to the appearance of *Kimberella* (Fedonkin & Waggoner, 1997). Molecular clock evidence puts the nearest common protostome-deuterostome ancestor between 600 and 1200 Mya (Reviewed in Erwin & Davidson, 2002). The sum of this evidence indicates that the existence of the protostome-deuterostome ancestor must predate 555 Mya (possibly quite significantly), and the divergence of Bilateria from Cnidaria most likely predates this by another 100My (Erwin & Davidson, 2002).

The Cambrian explosion, beginning 543 Mya, saw the rapid radiation of bilaterian fauna, with all major animal lineages suddenly appearing in the fossil record. Arthropods, annelids, molluscs and several chordates (in possession of segmented trunk muscles; Chen *et al.*, 2002) are all represented in Burgess Shale fauna, dated 525-515 Mya (Conway Morris, 1998). A fossil chelicerate has been identified from middle Cambrian deposits dated between 520-512 Mya (Briggs & Collins, 1988). Recognisable crustacean fossils appear in early Cambrian (more than 511 Mya; Siveter *et al.*, 2001). A phosphatocopid crustacean found in 511 Million year old rocks of Shropshire, England, has fully developed and segmented appendages (Siveter *et al.*, 2001). Insects are a more recent occurrence,

with the first insects appearing in the early Devonian (~400 Mya; Freeman and Herron, 2004). This dates the nearest common ancestor to both insects (*D. melanogaster*) and crustaceans (*C. destructor*) at greater than 511 Mya.

What this tells us is that segmented body plans, limbs, and the muscle used to make these two innovations function are observed in both protostomes and deuterostomes as soon as they are observed in the fossil record (Akam, 2000; Panganiban *et al.*, 1997). The inference of the morphological/developmental state of the nearest common ancestor of protostomes and deuterostomes is of great significance to our understanding of the role of *Pax* genes throughout evolution. Essentially, the question is whether *Pax* function is ancestral (plesiomorphic) or derived (apomorphic). If striated muscle were not present in the last common protostome-deuterostome ancestor, then any function of *pgIII* genes in myoblast specification must be derived, however, this does not seem to be the case.

4.5.3 The Evolution of Striated Muscle

Striated muscle is structurally and functionally conserved throughout higher Metazoa (triploblasts) (Cleto *et al.*, 2003; Spring *et al.*, 2002; Goodson & Spudich, 1993). All striated muscle is essentially comprised of the same structural proteins; myosin (heavy and light chains) and actin arranged into sarcomeres, with myosin providing the motor function (Bagshaw, 1993; Li & Mykles, 1990). Although myosins do exist in many organisms that lack striated muscle, comparisons of striated muscle-specific myosin in protostomes and deuterostomes indicates that both are clearly derived from a single common ancestor and that striated muscle predates the protostome-deuterostome split (Goodson & Spudich, 1993). Crustacean and vertebrate muscle also share a similar arrangement of tropomyosin and troponin-C, -T, and -I for Ca^{2+}-regulated contraction (Medler & Mykles, 2003; Mykles, 1988). Subtle differences do exist, for instance invertebrate sarcomeres differ in length between fast and slow fibres, and invertebrate thick filaments are bolstered with a paramyosin core (Mykles, 1988), but the essential characteristics that define striated muscle are conserved from vertebrates to arthropods.

Not only is striated muscle conserved throughout triploblastic animals (protostome and deuterostome lineages), but the same arrangement also occurs in certain species within the Cnidaria (which lack a true mesodermal germ layer) (Spring *et al.*, 2002). Because of the

high degree of ultrastructural and component protein (specifically myosin heavy chain and tropomyosin; Galliot & Schmid, 2002) similarity between cnidarian and triploblastic muscle, striated muscle must have arisen in the last common ancestor predating the Cnidaria-higher Metazoa split (Muller *et al.*, 2003). So, given that the structure of striated muscle is conserved between three animal phyla that are hugely disparate in body plan and separated by hundreds of millions of years of independent evolution, the question of interest in this research is whether the same regulatory mechanisms are used to produce striated muscle?

4.6 Molecular Regulation of Myogenesis

Embryonic myogenesis in vertebrates is accomplished by Wnt, Shh, and Bmp signalling in conjunction with regulation of the expression of *Pax3*, *Mef2* and *MyoD* (Amthor *et al.*, 1999). The emerging pattern for both embryonic and adult myogenesis in vertebrates is that inductive signals (Wnt/Shh/Bmp) act upstream of *pgIII* genes which in turn act upstream of the MyoD/Mef2 families of myogenic regulatory factors (Polesskaya *et al.*, 2003; Munsterberg *et al.*, 1995).

Wnt and Shh signals both induce myogenesis (Zorzano *et al.*, 2003; Munsterberg *et al.*, 1995). In the developing embryo, Wnt peptides are released from the neural tube, Shh protein is derived from the floor plate/notochord. These two molecules are sufficient to induce myogenesis in cultured somites (Munsterberg *et al.*, 1995). In contrast, Bmp is inhibitory of muscle differentiation. Bmp is released from the lateral plate mesoderm, and been shown to block *MyoD* expression (Pourquie *et al.*, 1996). Thus, myogenic patterning in the developing somite is a consequence of Wnt/Shh/Bmp gradients.

The MyoD family pathway and the Mef2 pathway regulate the differentiation of skeletal muscle in vertebrates (reviewed in Sabourin & Rudnicki, 2000). The MyoD family of transcription factors (MyoD, Myf5, MRF4 and myogenin) is responsible for the proliferation and differentiation required for the conversion of myoblasts into multinucleated muscle cells in both embryos and adult muscle (Sabourin & Rudnicki, 2000). The Mef2 family is also required for the determination and differentiation of both embryonic and adult myoblasts into striated muscle (reviewed in Dodou *et al.*, 2003).

By specifying myoblasts, *pgIII* genes create a vital link between patterning inductive signals (Wnt/Shh/Bmp) and myogenic determination and differentiation genes (*Mef2/MyoD*). *Pax3* is required for the specification of embryonic myoblasts, whereas *Pax7* is required for the specification of satellite cells (Seale *et al.*, 2000; Munsterberg *et al.*, 1995; Bober *et al.*, 1994). Expression of both *Pax3* and *Pax7* is induced by Wnt signals (Polesskaya *et al.*, 2003; Munsterberg *et al.*, 1995), and both *Pax3* and *Pax7* function as upstream regulators of the expression of the *MyoD* family and *Mef2* in myogenesis (Maroto *et al.*, 1997; Tajbakhsh, 2003; reviewed in Parker *et al.*, 2003). This indicates an important role for vertebrate *pgIII* genes in mediating the myogenic inductive effects of Wnt/Shh/Bmp pathways and the myogenic determination and differentiation MyoD/Mef2 pathways.

4.6.1 Myogenesis in *D. melanogaster*

The Wnt/Shh/Bmp inductive signalling pathways and the Mef2 and MyoD myogenic pathways are highly conserved. *D. melanogaster* has direct structural and functional homologues of Wnt (Wg; Wingless), Shh (Hh; Hedgehog) and Bmp (Dpp; Decapentaplegic) (Williams & Nagy, 2001; Jockusch *et al.*, 2000; Morata & Sanchez-Herrero, 1999; Reichmann *et al.*, 1997). These three inductive signals pattern the developing mesoderm into competence domains, from which muscle develops (Bidet *et al.*, 2003). Wg (Wnt) and Dpp (Bmp) also have antagonistic function in *D. melanogaster* as they do in vertebrates (Theisen *et al.*, 1996). There are, of course, differences in the embryonic development of arthropods and vertebrates. *D. melanogaster* does not have a notochord, so spatial orientation of the developing embryo is somewhat different.

D. melanogaster has only single *MyoD* and *Mef2* family members, whereas vertebrate *MyoD* and *Mef2* families are each represented by four genes (Dodou *et al.*, 2003). The *D. melanogaster MyoD* homologue *Nau* (*nautilus*) is expressed in embryonic myogenic progenitors (called muscle pioneer cells in *D. melanogaster*) and has function highly analogous to its vertebrate homologue in the determination and differentiation of striated muscle myoblasts (Zhang *et al.*, 1999). Null mutants for *Nau* have not yet been generated; however, evidence for this role comes from RNA interference studies, in which the

embryonic muscle patterning was severely disrupted when double-stranded *Nau* RNA was injected into embryos (Misquitta & Patterson, 1999). The function of *D. melanogaster* Mef2 is also highly analogous to that of the vertebrate Mef2 family transcription factors (Lin *et al.*, 1997; Dodou *et al.*, 2003). *D. melanogaster Mef2* null mutants show complete loss of muscle differentiation but no defects in myoblast specification (Lin *et al.*, 1997), indicating that, as in its vertebrate homologue, the *D. melanogaster Mef2* gene is involved in myogenic differentiation (Dodou *et al.*, 2003).

The role of Wg (Wnt) in *D. melanogaster* myogenic induction is highly analogous to that in vertebrates. *Wg* functions upstream of *Nau* and induces its expression in *D. melanogaster* muscle precursors, and *Wg* null mutants have a complete loss of *Nau* expressing medial muscle precursor cell clusters (Ranganayakulu *et al.*, 1996). Injection of either murine *Wnt1* or *D. melanogaster Wg* RNA into the frog *Xenopus laevis* results in identical biological effects (Ingham & Hidalgo, 1993) demonstrating functional equivalence of murine and insect Wnt homologues. Interestingly, Wg is the dominant activator of the *Gsb* late-enhancer (Deshpande *et al.*, 2001), indicating that the *D. melanogaster* Wnt homologue induces a *D. melanogaster pgIII* homologue. This Wg activated late expression of *Gsb* includes its expression in the developing mesoderm and in neural cells (Gutjahr *et al.*, 1993).

Wnt/Bmp and MyoD/Mef2 pathways are also present in Cnidaria, with analogous function (Muller *et al.*, 2003; Spring *et al.*, 2002). The hydrozoan Cnidarian *Podocoryne carnea* also possesses structural and functional homologues of higher metazoa MyoD and Mef2 transcription factors that act upstream of highly conserved muscle terminal differentiation markers (myosin heavy chain and tropomyosin) (Muller *et al.*, 2003; Spring *et al.*, 2002; Muller *et al.*, 1999). It is of interest that pgIII genes and striated muscle both first emerge in the Cnidaria, and that the Cnidaria also possess Wnt/Bmp and MyoD/Mef2 pathways.

D. melanogaster, like all other holometabolous insects, does not regenerate adult appendages (Slack, 2003). In hemimetabolous insects, however, limb regeneration is associated with Dpp/Hh/Wg secreting zones (reviewed in Slack, 2003). Cnidarians are also regenerators, and in *Hydra* the Wnt/Bmp pathways are markedly upregulated during early regeneration processes (Holstein *et al.*, 2003).

from *M. musculus* and *D. melanogaster* indicate conservation of myogenic inductive signalling (Wnt/Shh) and myogenic differentiation transcription factors (MyoD/Mef2). Morphological data from crustaceans and *M. musculus* indicate that the cellular processes involved in adult myogenesis are also analogous in an arthropod and a vertebrate. This picture is presented in Figure 4.5.

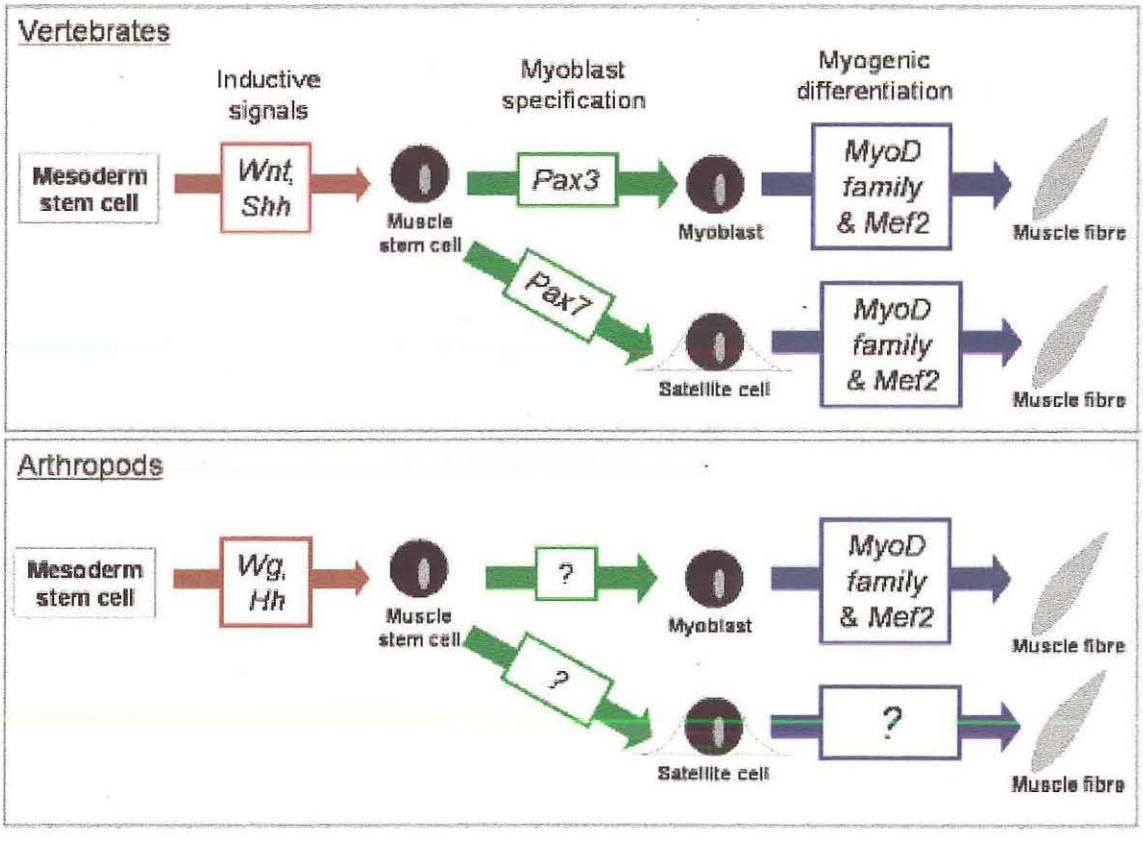


Figure 4.5: Cellular and genetic pathways involved in embryonic and adult myogenesis in vertebrates and arthropods. The vertebrate scenario is that described for *M. musculus* (Seale *et al.*, 2000; Goldring *et al.*, 2002; Ingham & McMahon, 2001; Zammit & Beauchamp, 2001; Bischoff, 1990; Goulding *et al.*, 1994). The arthropod scenario makes use of molecular data from *D. melanogaster* (Jockusch *et al.*, 2000; Misquitta & Patterson, 1999; Morata & Sanchez-Herrero, 1999; Reichmann *et al.*, 1997; Ranganayakulu *et al.*, 1996) and cellular data from decapod crustaceans including *C. destructor* (Cutler *et al.*, 2002; Lamey *et al.*, 2002; Read & Govind, 1998; Kao & Chang, 1997; Pearce *et al.*, 1997; Uhrik *et al.*, 1989 Erri-Babu, 1987). Question marks (?) indicate unknown pathways.

Given the conservation of the inductive mesodermal patterning regime, and the pathways of myogenic determination and differentiation in insects and vertebrates (and even jellyfish that lack a true mesoderm) it is fair to assume that these pathways also exist in Crustacea. *Wingless* (*Wnt*) homologues have been isolated in Crustacea, and have been demonstrated to be involved in the patterning of limbs during embryonic development (Williams & Nagy, 2001; Nulsen & Nagy, 1999).

The conservation of Wnt/Shh/Bmp and Mef2/MyoD pathways is certainly not predictive of a myogenic function for conserved *pgIII* genes. What it does tell us, however, is that there is a highly conserved framework for the induction of myogenesis into which *pgIII* gene action may fit. The expression of *Cdpax3/7* during *C. destructor* myogenesis may well fit into these pathways in a similar manner to vertebrate *pgIII* genes. Many muscle-specific genes are conserved throughout higher Metazoa, having probably originated prior to the Cnidaria-higher metazoa split and certainly prior to the divergence of protostomes and deuterostomes. *MyoD* and *Mef2* homologues are involved in myogenesis in Cnidarian, protostome and deuterostome striated muscle formation (Muller *et al.*, 2003; reviewed in Baylies & Michelson, 2001). The best explanation of the analogous expression patterns of *MyoD* and *Mef2* homologues in myogenesis throughout Cnidaria and higher Metazoa is that these homologues are under the control of analogous regulatory regions. In higher vertebrates, both *Pax3* and *Pax7* are important early regulators of myogenesis and act upstream of the *MyoD* family (Maroto *et al.*, 1997; Tajbakhsh, 2003; reviewed in Parker *et al.*, 2003).

So, does *C. destructor*, which has similar (though highly exacerbated) cellular patterns of muscle regeneration to vertebrates, use a *pgIII* gene to specify myoblasts? *Cdpax3/7* is expressed in muscle tissue containing quiescent (intermoult) and activated satellite cells (pre/postmoult), and is also expressed in regenerating tissue where adult stem cells are being redirected down the myogenic lineage, an expression pattern characteristic of *Pax7* in adult mice. It is thus tempting to speculate a role for *Cdpax3/7* that is analogous to *Pax3* and *Pax7* in the specification of myoblasts in higher vertebrates, and that this function may fit within the Wnt/Bmp and MyoD/Mef2 myogenic inductive pathways that are conserved from Cnidaria, to *D. melanogaster* to higher vertebrates.

4.7 Spanning the Great Divide: Functional Conservation amongst Bilateria?

4.7.1 Functional Evolution of *pgIII* genes

Implying conserved function of a transcription factor is essentially stating that there is evolutionary linkage between the transcription factor and promoter regions of the genes it activates. It is apparent that three criteria are essential in this logic; (1) homologous transcription factors must bind to similar recognition sequences, (2) similar recognition sequences must occur in the promoter regions of downstream target genes, and (3) the transcription factor must exhibit similar patterns of expression throughout evolution.

The sequence of pgIII transcription factors is highly conserved between Cdpax3/7, and its protostome and deuterostome homologues. As the previous analysis of the DNA-binding properties of these transcription factors has indicated, Cdpax3/7 contains many of the same amino acids that have been identified to be functionally important in vertebrate Pax3 (Apuzzo & Gros, 2002; Underhill & Gros, 1997). This gives strong support to the idea that Cdpax3/7 may bind similar DNA recognition sequences to vertebrate Pax3 and Pax7.

The second criterion is whether the promoter regions of muscle tissue specific genes are conserved. Many of the striated muscle-specific transcriptional hierarchies that act downstream of *Pax3* and *Pax7* in higher vertebrates are highly structurally and functionally conserved throughout Bilateria (Zhang *et al.*, 1999; Baylies & Michelson, 2001), and evidently had their origins prior to the Cnidaria/higher Metazoa split (Spring *et al.*, 2002). The probability that these hierarchical expression patterns are conserved due to the promoter regions of these muscle-specific genes being conserved far outweighs the possibility that co-option of each of these genes has occurred independently.

This research provides some support towards satisfaction of the third criterion. *Cdpax3/7* is expressed in normal adult muscle, muscle undergoing regeneration, and also in pre-muscle tissue undergoing myogenesis (the conversion of non-muscle stem cells into muscle). This gross pattern of expression is analogous between *Cdpax3/7* and its vertebrate *Pax7* homologue. *Cdpax3/7* is also expressed concurrently with limb formation in the developing embryo, as are vertebrate *Pax3* and *Pax7*. However, expression in the embryo

at the same time as limb formation is certainly not sufficient to imply any role in this process, for instance, CNS patterning would also be occurring in the embryo at this stage, and it is likely that *Cdpax3/7* would also play some role in CNS patterning like its *pgIII* homologues in other species (Dearden *et al.*, 2002; Davis *et al.*, 2001; Holland *et al.*, 1999; Gutjahr *et al.*, 1993; Jostes *et al.*, 1991). Studies of tissue-specific expression patterns of *Cdpax3/7* during embryogenesis are necessary to develop this idea; however this must be preceded by a descriptive study of freshwater crayfish embryology and tissue patterning including labelling for muscle-specific proteins (such as MyoD and Mef2 homologues).

The question must therefore be asked, when did the specification of myoblasts by *pgIII* genes arise? *Pax3* and *Pax7* expression patterns appear to be conserved throughout mammals, birds and fish (Seo *et al.*, 1998; Jostes *et al.*, 1991). Studies in amphioxus (Cephalochordata) indicate that *Amphipax3/7* (the nearest common ancestral progenitor to both *Pax3* and *Pax7*) is involved in myogenesis, indicating a role analogous to *Pax3* and *Pax7* in specifying myoblast lineages (Holland *et al.*, 1999). *Amphipax3/7* is expressed in the somites of the developing mesoderm and throughout the axial musculature (Holland *et al.*, 1999). The research presented here is the first report of expression of a *pgIII* gene during myogenesis outside of the phylum Chordata.

4.7.2 Gene Duplication and Its Role in Genetic Functional Diversification

It is widely accepted that the dominant cause of change in gene regulation (regulatory regions/expression patterns/functions) is gene duplication (Seoighe *et al.*, 2003; Force *et al.*, 1999; Holland, 2003; Mayr, 2002; Frazer *et al.*, 2003). After duplication, genes are somewhat released from the constraints of normalising selective pressure, and thus mutations resulting in the functional changes may occur (Force *et al.*, 2003; Xue & Noll, 1996; Li & Noll, 1994).

The *pgIII* evolutionary scenario best elucidated to date places *Cdpax3/7* and *Amphipax3/7* as direct orthologues (single genes derived from a common ancestor by speciation), not paralogues that have arisen via duplication. So why would these two genes not be in possession of similar regulatory elements? By this approximation, it is the insect *pgIII* genes (that have undergone extensive duplication) that would be expected to have divergent

function. None of the *D. melanogaster pgIII* expression patterns can account for the observed expression of *Cdpax3/7* in tissues undergoing muscle regeneration, while the amphioxus/vertebrate *pgIII* gene expression pattern can.

5.0 CONCLUSIONS, MAJOR FINDINGS AND FURTHER RESEARCH

5.1 Conclusions and Major Findings

The evolution of developmental transcription factors such as the *Pax* family of genes is pivotal to our understanding of the diversity of forms and functions present throughout the animal kingdom. This research provides the first description of a *Pax* gene in Crustacea. *C. destructor* has a single *pgIII* gene, *Cdpax3/7*. The deduced Cdpax3/7 protein contains all the hallmarks of a pgIII protein; highly conserved putative DNA binding domains (the HD and the bipartite PD), as well as an octapeptide motif in the PD/HD linker region. Phylogenetic analysis positions *Cdpax3/7* in an ancestral location to all *pgIII* genes isolated in insects, indicating that the progenitor of *Cdpax3/7* has given rise to *Pby1* and *Pby2* in *S. americana* and *Prd*, *Gsb*, and *Gsb-n* in *D. melanogaster* via gene duplication.

Many parallels exist between the structure of *Cdpax3/7* and its vertebrate homologues *Pax3* and *Pax7*. High homology of DNA binding domains and an almost exact identity of amino acids at critical DNA contacting positions indicate that Cdpax3/7 may display similar DNA binding properties to the vertebrate proteins, and through this may have similar function.

Remarkably, Cdpax3/7 and vertebrate Pax3 and Pax7 share a common splice site at position 75 of the PD, that is not conserved in any other *pgIII* genes isolated to date. Alternate splicing (the addition of a glutamine residue) at this position is critical in modulating the function of Pax3 and Pax7 proteins. Interestingly, *Cdpax3/7* uses this splice site to produce a somewhat more drastically modified alternate transcript, lacking 93 nucleotides encoding the first two helices of the RED domain of the PD. The other splice site identified in *Cdpax3/7* at position 105 of the PD is conserved in one of the *S. americana* genes (*Pby1*) and as an intron in one of the *D. melanogaster* genes (*Gsb-n*). Taken together these data imply that (1) the position 105 splice site appears to be an ancestral arthropod *pgIII* relic, (2) that the position 75 splice site appears to be an ancestral bilaterian *pgIII* relic, and (3) that the *Cdpax3/7-b* alternate transcript appears to be generated by exon skipping from these two putative intron sites.

It is likely that this alternate transcript modifies the DNA binding functions of the RED domain of the PD of Cdpax3/7. Thus the *Cdpax3/7* gene produces two proteins, one which may bind DNA using its entire PD and HD, the other which may bind DNA utilising only the PAI domain of its PD coupled with its HD. This may allow a single *pgIII* gene present in *C. destructor* to satisfy the multitude of functions performed by multiple *pgIII* genes in other animals.

This research marks the tip of the iceberg of the genetic mechanisms underlying the complex and astonishing process of muscle regeneration in the decapod crustacean *C. destructor*. *Cdpax3/7* is expressed during adult muscle regeneration in the claw closer muscle of *C. destructor*. *Cdpax3/7* is expressed during normal muscle regeneration processes (premoult, postmoult and intermoult) and also during extreme (epimorphic) muscle regeneration, both before and after myogenic differentiation. Thus the gross expression pattern of *Cdpax3/7* during adult muscle regeneration parallels that of vertebrate *Pax7* in satellite cells and the conversion of SP cells into satellite cells. Myogenic pathways identified to act upstream (Wnt/Shh) and downstream (MyoD/Mef2) of *pgIII* genes in vertebrate myogenesis, are highly functionally conserved throughout animal phyla. Thus further research into the interactions between *Cdpax3/7* expression during myogenesis with crustacean homologues of *Wnt/Shh* and *MyoD/Mef2* may unearth another link in this highly conserved hierarchy. Should these myogenic pathways turn out to be conserved between Crustacea and vertebrates, the crustacean model of limitless adult myogenesis may give important insights into how we may treat degenerative muscular diseases in vertebrates. The inexhaustible adult muscle regeneration observed in *C. destructor* is a phenomenon that is certainly worth researching in this regard.

5.2 Further Research

To further elucidate the protostome evolutionary lineage of *pgIII* genes, specifically pinning down the duplications that have given rise to the three *D. melanogaster* and two *S. americana* genes, it will be necessary to isolate *pgIII* genes from taxa more closely related to Diptera and Orthoptera. This may also help trace the evolutionary history of the splice site at position 75 of the PD conserved between *Cdpax3/7* and vertebrate *Pax3* and *Pax7*.

Characterising the expression patterns of *pgIII* genes in other adult invertebrates may help to establish a conserved role for *pgIII* genes in regeneration, to add to their known conserved roles in embryonic development.

To establish a role for *Cdpax3/7* in the specification of myoblasts, it will be necessary to localise *Cdpax3/7* expression to specific cells in regenerating muscle and epimorphically regenerating limb samples from *C. destructor* using *in situ* hybridisation. After localisation of *Cdpax3/7* transcripts to a particular cell lineage, studying the effects of misexpression by utilising double stranded *Cdpax3/7* RNA interference techniques, or overexpression by the insertion of a *Cdpax3/7* carrying expression vector, will help elucidate a functional role for *Cdpax3/7* in adult muscle regeneration of *C. destructor*. The generation of *Cdpax3/7* mutants would also give a good indication of the summed functions of this gene.

Myogenic pathways need to be determined for crustaceans. Members of the *MyoD* and *Mef2* families are literally waiting to be found in Crustacea. Following this, an analysis of the relative expression patterns of *Cdpax3/7*, *MyoD/Mef2*, and *Wg* are required to help discern specific myogenic hierarchies in Crustacea.

It will be important to study the expression patterns and functions of *Cdpax3/7* during embryonic development in *C. destructor*. Determining the expression patterns and functions of *Cdpax3/7* during embryonic segmentation, neurogenesis, and myogenesis will generate an important understanding of both the functions of *Cdpax3/7* and of the functional evolution of transcription factor genes.

Characterisation of the genomic organisation of the *Cdpax3/7* gene is necessary to compare intronic sequences with those of homologous *pgIII* genes. Of particular interest is the identification of the intron/splice site at amino acid position 75 of the PD, which is conserved in vertebrate *Pax3* and *Pax7*.

An analysis of the DNA sequence binding specificities of Cdpax3/7 should be conducted using electrophoretic mobility shift assays. This should be conducted using both Cdpax3/7-a and Cdpax3/7-b proteins, as it is probable that these two isoforms will recognise different DNA sequences.

6.0 REFERENCES

Agosti, D., Jacobs, D., & DeSalle, R. (1996). On combining protein sequences and nucleic acid sequences in phylogenetic analysis: The homeobox protein case. *Cladistics, 12*, 65-82.

Akam, M. (2000). Arthropods: Developmental diversity within a (super) phylum. *Proceedings of the National Academy of Sciences of the United States of America, 97*(9), 4438-4441.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein-database search programs. *Nucleic Acids Research, 25*(17), 3389-3402.

Apuzzo, S., & Gros, P. (2002). Site-specific modification of single cysteine Pax 3 mutants reveals reciprocal regulation of DNA binding activity of the paired and homeo domain. *Biochemistry, 41*, 12076-12085.

Asakura, A., Komaki, M., & Rudnicki, M. A. (2001). Muscle satellite cells are multipotential stem cells that exhibit myogenic, osteogenic, and adipogenic differentiation. *Differentiation, 68*, 245-253.

Asakura, A., Seale, P., Girgis-Gabardo, A., & Rudnicki, M. A. (2002). Myogenic specification of side population cells in skeletal muscles. *Journal of Cell Biology, 159*(1), 123-134.

Bagshaw, C. R. (1993). *Muscle contraction* (2nd ed.). London: Chapman & Hall.

Barr, F. G., Fitzgerald, J. C., Ginsberg, J. P., Vanella, M. L., Davis, R. J., & Bennicelli, J. L. (1999). Predominant expression of alternative PAX3 and PAX7 forms in myogenic and neural tumor cell lines. *Cancer Research, 59*, 5443-5448.

Balczarek, K. A., Lai, Z.-C., & Kumar, S. (1997). Evolution and functional diversification of the paired box (*Pax*) DNA-binding domains. *Molecular Biology and Evolution, 14*(8), 829-842.

Baumgartner, S., Bopp, D., Burri, M., & Noll, M. (1987). Structure of two genes at the *gooseberry* locus related to the *paired* gene and their spatial expression during *Drosophila* embryogenesis. *Genes and Development, 1*, 1247-1267.

Baylies, M. K., & Michelson, A. M. (2001). Invertebrate myogenesis: looking back to the future of muscle development. *Current Opinion in Genetics & Development, 11*, 431-439.

Bertuccioli, C., Fasano, L., Jun, S., Wang, S., Sheng, G., & Desplan, C. (1996). In vivo requirement for the paired domain and homeodomain of the *paired* segmentation gene product. *Development, 122*, 2673-2685.

Bidet, Y., Jagla, T., Da Ponte, J.-P., Dastugue, B., & Jagla, K. (2003). Modifiers of muscle and heart cell fate specification identified by gain-of-function screen in *Drosophila*. *Mechanisms of Development, 120*, 991-1007.

Bischoff, R. (1990). Cell cycle commitment of rat muscle satellite cells. *Journal of Cell Biology, 111*, 201-207.

Bober, E., Franz, T., Arnold, H.-H., Gruss, P., & Tremblay, P. (1994). *Pax-3* is required for the development of limb muscles: a possible role for the migration of dermomyotomal muscle progenitor cells. *Development, 120*, 603-612.

Borycki, A.-G., Li, J., Jin, F., Emerson, C. P., & Epstein, J. A. (1999). Pax3 functions in cell survival and in *pax7* regulation. *Development, 126*, 1665-1674.

Bouchard, M., St-Amand, J., & Cote, S. (2000). Combinatorial activity of pair-rule proteins on the *Drosophila gooseberry* early enhancer. *Developmental Biology, 222*, 135-146.

Breitling, R., & Gerber, J.-K. (2000). Origin of the paired domain. *Developmental Genetics and Evolution, 210*, 644-650.

Briggs, D. E. G., & Collins, D. (1988). A middle Cambrian chelicerate from Mount Stephens, British Columbia. *Palaeontology, 31*, 779-798.

Burton, E. M., & Mitchell, B. D. (1987). Moult staging in the Australian freshwater crayfish, *Cherax albidus* Clark and *Cherax destructor* Clark (Decapoda: Parastacidae), via uropod setal development. *Australian Journal of Marine and Freshwater Research, 38*, 545-552.

Campos-Ortega, J. A., & Hartenstein, V. (1985). *The embryonic development of Drosophila melanogaster*. Berlin: Springer-Verlag.

Carlson, B. M. (2003). Muscle regeneration in amphibians and mammals: Passing the torch. *Developmental Dynamics, 226*, 167-181.

Chalepakis, G., & Gruss, P. (1995). Identification of DNA recognition sequences for the Pax3 paired domain. *Gene, 162*, 267-270.

Chen, J.-Y., Oliveri, P., Gao, F., Dornbos, S. Q., Li, C.-W., Bottjer, D. J., & Davidson, E. H. (2002). Precambrian animal life: probable developmental and adult cnidarian forms from Southwest China. *Developmental Biology, 28*(1),182-196.

Christ, B., & Ordahl, C. P. (1995). Early stages of chick somite development. *Anatomy and Embryology, 191*, 381-396.

Chromas version 2.3 [Computer software]. (2002). Queensland, Australia: Technelysium Pty. Ltd.

Cleto, C. L., Vandenberghe, A. E., MacLean, D. W., Pannunzio, P., Tortorelli, C., Meedel, T. H., Satou, Y., Satoh, N., & Hastings, K. E. M. (2003). Ascidian larva reveals ancient origin of vertebrate-skeletal-muscle troponin I characteristics in chordate locomotory muscle. *Molecular Biology and Evolution.* (in press).

Conway Morris, S. (1998). Early metozoan evolution: Reconciling palaeontology and molecular biology. *American Zoologist, 38*, 867-877.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research,16*(22), 10881-10890.

Cutler, L. K., Koenders, A., Klemm., M. F., West, J. M., & Mykles, D. L. (2002). Myofibrillar protein composition of muscle fibres from regenerating and pristine claws of the freshwater crayfish *Cherax destructor.* In G. J. Whisson, & B. Knott (Eds.), *Freshwater crayfish 13: Proceedings of the thirteenth symposium of the International Association of Astacology* (pp. 479-490). Perth: Curtin University.

Czerny, T., Schaffner, G., & Busslinger, M. (1993). DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site. *Genes & Development, 7*, 2048-2061.

Davis, G. A., & Patel, N. H. (2002). Short, long, and beyond: Molecular and embryological approaches to insect segmentation. *Annual Review of Entomology, 47*, 669-699.

Davis, G. K., & Patel, N. H. (1999). The origin and evolution of segmentation. *Trends in Genetics, 15*(12), M68-72.

Davis, G. K., Jaramillo, C. A., & Patel, N. H. (2001). Pax group III genes and the evolution of insect pair-rule patterning. *Development, 128*, 3445-3458.

Dearden, P. K., Donly, C., & Grbic, M. (2002). Expression of pair-rule gene homologues in a chelicerate: early patterning of the two-spotted mite *Tetranychus urticae. Development, 129*, 5461-5472.

Deshpande, N., Dittrich, R., Technau, G. M., & Urban, J. (2001). Successive specification of *Drosophila* neuroblasts NB 6-4 and NB 7-3 depends on interaction of the segment polarity genes *wingless, gooseberry* and *naked cuticle. Development, 128*, 3253-3261.

Dodou, E., Xu, S.-M., & Black, B. L. (2003). *mef2c* is activated directly by myogenic basic helix-loop-helix proteins during skeletal muscle development in vivo. *Mechanisms of Development, 120*, 1021-1032.

Duman-Scheel, M., Li, X., Orlov, I., Noll, M., & Patel, N. H. (1997). Genetic separation of the neural and cuticular patterning functions of *gooseberry. Development, 124*, 2855-2865.

Erri Babu, D. (1987). Histological and histochemical studies on regeneration and tissue differentiation in the crab *Menippe rumphii* (Fabricius) (Crustacea: Brachyura). *Journal Experimental Marine Biology and Ecology, 111*, 213-230.

Erwin, D. H., & Davidson, E. H. (2002). The last common bilaterian ancestor. *Development, 129*, 3021-3032.

Fedonkin, M. A., & Waggoner, B. M. (1997). The late Precambrian fossil *Kimberella* is a mollusc-like bilaterian organism. *Nature, 388*, 868-871.

Felsenstein, J., & Churchill, G. (1996). A hidden Markov model approach to variation among sites in rate of evolution. *Molecular Biology and Evolution, 13*, 93-104.

Felsenstein, J., & Sober, E. (1986). Parsimony and likelihood: an exchange. *Systematic Zoology, 35*, 617-626.

Ferrari, G., Cusella-De Angelis, G., Coletta, M., Paolucci, E., Stornaiuolo, A., Cossu, G., & Mavilio, F. (1998). Muscle regeneration by bone marrow-derived myogenic precursors. *Science, 279*,1528-1530.

Ferrari, G., Stornaiuolo, A., & Mavilio, F. (2001). Failure to correct murine muscular dystrophy. *Nature*, 411, 1014-1015.

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

Francis-West, P. H., Antoni, L., & Anakwe, K. (2003). Regulation of myogenic differentiation in the developing limb bud. *Journal of Anatomy, 202*, 69-81.

Frazer, K. A., Elnitski, L., Church, D. M., Dubchak, I., & Hardison, R. C. (2003). Cross-species sequence comparisons: A review of methods and available resources. *Genome Research, 13*, 1-12.

Freeman, S., & Herron, J. C. (2004). *Evolutionary Analysis* (3rd ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S., & Noll, M. (1986). Structure of the segmentation gene paired and the Drosophila PRD gene set as part of a gene network. *Cell, 47*(5), 735-746.

Galliot, B., & Schmid, V. (2002). Cnidarians as a model system for understanding evolution and regeneration. *International Journal of Developmental Biology, 46*, 39-48.

Goldring, K., Partridge, T. A., & Watt, D. (2002). Muscle stem cells. *Journal of Pathology, 197*, 457-467.

Goodson, H. V., & Spudich, J. A. (1993). Molecular evolution of the myosin family: Relationships derived from comparisons of amino acid sequences. *Proceedings of the National Academy of Sciences of the United States of America, 90*, 659-663.

Goulding, M., Lumsden, A., & Paquette, A. J. (1994). Regulation of *Pax-3* expression in the dermomyotome and its role in muscle development. *Development, 120*, 957-971.

Govind, C. K., & Pearce, J. (1994). Muscle remodelling in adult snapping shrimps via fast-fibre degeneration and slow-fibre genesis and transformation. *Cell and Tissue Research, 276*, 445-454.

Groger, H., Callaerts, P., Gehring, W. J., & Schmid, V. (2000). Characterization and expression analysis of an ancestor-type *Pax* gene in the hydrozoan jellyfish *Podocoryne carnea. Mechanisms of Development, 94*, 157-169.

Grounds, M. D., White, J. D., Rosenthal, N., & Bogoyevitch, M. A. (2002). The role of stem cells in skeletal and cardiac muscle repair. *Journal of Histochemistry and Cytochemistry, 50*(5), 589-610.

Grounds, M. D., & Yablonka-Reuveni, Z. (1993). Molecular and cell biology of skeletal muscle regeneration. In T. Partridge, (Ed.), *Molecular and cell biology of muscular dystrophy* (pp. 210-255). London: Chapman & Hall.

Gussoni, E., Soneoka, Y., Strickland, C. D., Buzney, E. A., Khan, M. K., Flint, A. F., Kunkel, L. M., & Mulligan, R. C. (1999). Dystrophin expression in the *mdx* mouse restored by stem cell transplantation. *Nature, 401*, 390-394.

Gutjahr, T., Patel, N. H., Li, X., Goodman, C. S., & Noll, M. (1993). Analysis of the *gooseberry* locus in *Drosophila* embryos: *gooseberry* determines the cuticular pattern and activates *gooseberry neuro. Development, 118*, 21-31.

Halder, G., Callaerts, P., & Gehring, W. (1995). Induction of ectopic eyes by targeted expression of the *eyeless* gene in *Drosophila. Science, 267*, 1788-1792.

Hamr, P. (1992). Embryonic and postembryonic development in the Tasmanian freshwater crayfishes *Astacopsis gouldi, Astacopsis franklinii* and *Parastacoides tasmanicus tasmanicus* (Decapoda: Parastacidae). *Australian Journal of Marine and Freshwater Research, 43*, 861-878.

Hawke, T. J., & Garry, D. J. (2001). Myogenic satellite cells: physiology to molecular biology. *Journal of Applied Physiology, 91*, 534-551.

Higgins, D., Thompson, J., Gibson, T., Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research, 22*, 4673-4680.

Holland, C. A., & Skinner, D. M. (1976). Interactions between moulting and regeneration in the land crab. *Biological Bulletin, 150*, 222-240.

Holland, L. Z., Schubert, M., Kozmik, Z., & Holland, N. D. (1999). *AmphiPax3/7*, an amphioxus paired box gene: insights into chordate myogenesis, neurogenesis, and the possible evolutionary precursor of definitive vertebrate neural crest. *Evolution & Development, 1*(3), 153-165.

Holland, N. D., Holland, L. Z., & Kozmik, Z. (1995). An amphioxus *Pax* gene, *AmphiPax-1*, expressed in the embryonic endoderm, but not in the mesoderm: Implications for the evolution of class I paired box genes. *Molecular Marine Biology and Biotechnology, 4*, 206-214.

Holland, P. W. H. (2003). More genes in vertebrates? *Journal of Structural and Functional Genomics, 3*, 75-84.

Holstein, T. W., Hobmayer, E., & Tachnau, U. (2003). Cnidarians: An evolutionarily conserved model system for regeneration? *Developmental Dynamics, 226*, 257-267.

Hopkins, P. M., Chung, A. C.-K., & Durica, D. S. (1999). Limb regeneration in the fiddler crab, *Uca pugilator*: Histological, physiological and molecular considerations. *American Zoology, 39*, 513-526.

Ingham, P. W., & Hidalgo, A. (1993). Regulation of *wingless* transcription in the *Drosophila* embryo. *Development, 117*, 283-291.

Ingham, P. W., & McMahon, A. P. (2001). Hedgehog signaling in animal development: paradigms and principles. *Genes & Development, 15*, 3059-3087.

Jankowski, R. J., Deasy, B. M., Cao, B., Gates, C., & Huard, J. (2002). The role of CD34 expression and cellular fusion in the regeneration capacity of myogenic progenitor cells. *Journal of Cell Science, 115*, 4361-4374.

Jockusch, E. L., Nulsen, C., Newfeld, S. J., & Nagy, L. M. (2000). Leg development in flies versus grasshoppers: differences in *dpp* expression do not lead to differences in the expression of downstream components of the leg patterning pathway. *Development, 127*, 1617-1626.

Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences, 8*, 275-282.

Jostes, B., Walther, C., & Gruss, P. (1991). The murine paired box gene, *Pax7* is expressed specifically during the development of the nervous and muscular system. *Mechanisms of Development, 33*, 27-37.

Jun, S., & Desplan, C. (1996). Cooperative interactions between paired domain and homeodomain. *Development, 122*, 2639-2650.

Jun, S., Wallen, R. V., Goriely, A., Kalionis, B., & Desplan, C. (1998). Lune/eye gone, a Pax-like protein, uses a partial paired domain and a homeodomain for DNA recognition. *Proceedings of the National Academy of Sciences of the United States of America, 95*, 13720-13725.

Kao, H.-W., & Chang, E. S. (1997). Limb regeneration in the eye sockets of crabs. *Biological Bulletin, 193*, 393-400.

Kilchherr, F., Baumgartner, S., Bopp, D., Frei, E., & Noll, M. (1986). Isolation of the *paired* gene of *Drosophila* and its spatial expression during early embryogenesis. *Nature, 321*, 493-499.

Kos, L., Chiang, C., Mahon, K. A. (1998). Mediolateral patterning of somites: multiple axial signals, including *Sonic hedgehog*, regulate *Nkx*-3.1 expression. *Mechanisms of Development, 70*, 25-34.

Kozmik, Z., Czerny, T., & Busslinger, M. (1997). Alternatively spliced insertions in the paired domain restrict the DNA sequence specificity of Pax6 and Pax8. *European Molecular Biology Organisation Journal, 16*(22), 6793-6803.

Kozmik, Z., Holland, N. D., Kalusova, A., Paces, J., Schubert, M., & Holland, L. Z. (1999). Characterisation of an amphioxus paired box gene *AmphiPax2/5/8*: Developmental expression patterns in the optic support cells, nephridium, thyroid-like structures and pharyngeal gill slits, but not in the midbrain-hindbrain region. *Development, 126*, 1295-1304.

LaBarge, M. A., & Blau, H. M. (2002). Biological progression from adult bone marrow to mononucleate muscle stem cell to multinucleate muscle fiber in response to injury. *Cell, 111*, 589-601.

Lamey, T., Koenders, A., & West, J. M. (2002). The effects of moult-induced atrophy on claw muscle of the yabby, *Cherax destructor*. In G. J. Whisson, & B. Knott (Eds.), *Freshwater crayfish 13: Proceedings of the thirteenth symposium of the International Association of Astacology* (pp. 479-490). Perth: Curtin University.

Lechner, M. S., & Dressler, G. R. (1996). Mapping of Pax-2 transcription activation domains. *Journal of Biological Chemistry, 271*(35), 21088-21093.

Li, C.-W., Chen, J.-Y., & Hua, T.-E. (1998). Precambrian sponges with cellular structures. *Science, 279*, 879-882.

Li, X., & Noll, M. (1994). Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different *cis*-regulatory regions. *Nature, 367*, 83-87.

Li, Y., & Mykles, D. L. (1990). Analysis of myosins from lobster muscles: fast and slow isozymes differ in heavy-chain composition. *Journal of Experimental Zoology, 207*, 163-170.

Lin, S.-C., Lin, M.-H., Horvath, P., Reddy, K. L., & Storti, R. V. (1997). PDP1, a novel *Drosophila* PAR domain bZIP transcription factor expressed in developing mesoderm, endoderm and ectoderm, is a transcriptional regulator of somatic muscle genes. *Development, 124*, 4685-4696.

Maroto, M., Reshef, R., Munsterberg, A. E., Koester, S., Goulding, M., & Lassar, A. B. (1997). Ectopic *Pax-3* activates *MyoD* and *Myf-5* expression in embryonic mesoderm and neural tissue. *Cell, 89*, 139-148.

Martin, M. W., Grazhdankin, D. V., Bowring, S. A., Evans, D. A., Fedonkin, M. A., & Kirschvink, J. L. (2000). Age of Neoproterozoic bilaterian body and trace fossils, White Sea, Russia: implications for metazoan evolution. *Science, 288*, 841-845.

Mauro, A. (1961). Satellite cell of muscle fibres. *Journal of Biophysical and Biochemical Cytology, 99*, 493-498.

Mayr, E. (2002). *What evolution is*. London: Phoenix.

McKinney-Freeman, S. L., Jackson, K. A., Camargo, F. D., Ferrari, G., Mavilio, F., & Goodell, M. A. (2002). Muscle-derived hematopoietic stem cells are hematopoietic in origin. *Proceedings of the National Academy of Sciences of the United States of America, 99*(3), 1341-1346.

Medler, S., & Mykles, D. L. (2003). Analysis of myofibrillar proteins and transcripts in adult skeletal muscles of the American lobster *Homarus americanus*: variable expression of myosins, actin and troponins in fast, slow-twitch and slow-tonic fibres. *Journal of Experimental Biology, 206*, 3557-3567.

Michelson, A. M. (1994). Muscle pattern diversification in *Drosophila* is determined by the autonomous function of homeotic genes in the embryonic mesoderm. *Development, 120*, 755-768.

Miller, D. J., Hayward, D. C., Reece-Hoyes, J. S., Scholten, I., Catmull, J , Gehring, W., Callaerts, P., Larsen, J. E., & Ball, E. E. (2000). *Pax* gene diversity in the basal cnidarian *Acropora millepora* (Cnidaria, Anthozoa): Implications for the evolution of the *Pax* gene family. *Proceedings of the National Academy of Sciences of the United States of America, 97*(9), 4475-4480.

Miriami, E., Margalit, H., & Sperling, R. (2003). Conserved sequence elements associated with exon skipping. *Nucleic Acids Research, 31*(7), 1974-1983.

Miskiewicz, P., Morrissey, D., Lan, Y., Raj, L., Kessler, S., Fujioka, M., Goto, T., & Weir, M. (1996). Both the paired domain and homeodomain are required for in vivo function of *Drosophila* Paired. *Development, 122* 2709-2718.

Misquitta, L., & Paterson, B. M. (1999). Targeted disruption of gene function in *Drosophila* by RNA interference (RNA-i): A role for *nautilus* in embryonic somatic muscle formation. *Proceedings of the National Academy of Sciences of the United States of America, 96*, 1451-1456.

Morata, G., & Sanchez-Herrero, E. (1999). Patterning mechanisms in the body trunk and the appendages of *Drosophila*. *Development, 126*, 2823-2828.

Muller, P., Seipel, K., Yanze, N., Reber-Muller, S., Streitwolf-Engel, R., Steirwald, M., Spring, J., & Schmid, V. (2003). Evolutionary aspects of developmentally regulated helix-loop-helix transcription factors in striated muscle of jellyfish. *Developmental Biology, 255*, 216-229.

Muller, P., Yanze, N., Schmid, V., & Spring, J. (1999). The homeobox gene *Otx* of the jellyfish *Podocoryne carnea*: Role of a head gene in striated muscle and evolution. *Developmental Biology, 216*, 582-594.

Munsterberg, A. E., Kitajewski, J., Bumcrot, D. A., McMahon, A. P., & Lassar, A. B. (1995). Combinatorial signalling by Sonic hedgehog and Wnt family members induces myogenic bHLH gene expression in the somite. *Genes Development, 9*, 2911-2922.

Mykles, D. L., & Skinner, D. M. (1982). Crustacean muscles: atrophy and regeneration during molting. In B. M. Twarog, R. J. C. Levine & M. M. Dewey (Eds.), *Basic biology of muscles: a comparative approach* (pp. 337-357). New York: Raven Press.

Mykles, D. L. (1988). Histochemical and biochemical characterization of two slow fiber types in decapod crustacean muscles. *Journal of Experimental Zoology, 245*, 232-243.

Noll, M. (1993). The evolution and role of *Pax* genes. *Current Opinions in Genetics & Development, 3*, 595-605.

Northwestern University. *Oligo properties calculator version 3.02* (Computer software). (2002). Chicago, Illinois.

Novotova, M., & Uhrik, B. (1992). Structural characteristics and distribution of satellite cells along crayfish muscle fibers. *Experientia, 48*, 593-596.

Nulsen, C., & Nagy, L. M. (1999). The role of wingless in the development of multibranched crustacean limbs. *Development Genes and Evolution, 209*(6), 340-348.

Page, R. D. M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**: 357-358.

Panganiban, G., Irvine, S. M., Lowe, C., Roehl, H., Corley, L. S., Sherbon, B., Grenier, J. K., Fallon, J. F., Kimble, J., Walker, M., Wray, G. A., Swalla, B. J., Martindale, M. Q., & Carroll, S. B. (1997). The origin and evolution of animal appendages. *Proceedings of the National Academy of Sciences of the United States of America, 94*, 5162-5166.

Parker, M. H., Seale, P., & Rudnicki, M. A. (2003). Looking back to the embryo: Defining transcriptional networks in adult myogenesis. *Nature Reviews Genetics, 4*, 497-507.

Pearce, J., Krause, K. M., & Govind, C. K. (1997). Muscle fibres in regenerating crayfish motor nerves. *Journal of Neurophysiology, 78*(6), 3498-3501.

Pearson, W. R. (1990). Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods in Enzymology, 183*, 63-98.

PHYLIP (Phylogeny inference package version 3.6a3) [computer software]. (2002). Seattle: Department of Genome Sciences, University of Washington.

Polesskaya, A., Seale, P. & Rudnicki, M. A. (2003). Wnt signalling induces the myogenic specification of resident $CD45^+$ adult stem cells during muscle regeneration. *Cell, 113*, 841-852.

Pourquie, O., Fan, C.-M., Coltey, M., Hirsinger, E., Watanabe, Y., Breant, C., Francis-West, P., Brickell, P., Tessier-Lavigne, M., & Le Douarin, N. M. (1996). Lateral and axial signals involved in avian somite patterning: a role for BMP4. *Cell, 84*, 461-471.

Qiagen. (2002). *Qiagen OneStep RT-PCR kit handbook.* (Booklet). California: United States of America.

Qu-Petersen, Z., Deasy, B., Jankowski, R., Ikezawa, M., Cummins, J., Pruchnic, R., Mytinger, J., Cao, B., Gates, C., Wernig, A., & Huard, J. (2002). Identification of a novel population of muscle stem cells in mice: potential for muscle regeneration. *Journal of Cell Biology, 157*(5), 851-864.

Ranganayakulu, G., Schulz, R. A., & Olson, E. N. (1996). *Wingless* signaling induces *nautilus* expression in the ventral mesoderm of the *Drosophila* embryo. *Developmental Biology, 176*(1), 143-148.

Read, A. T., & Govind, C. K. (1998). Cell types in regenerating claws of the snapping shrimp, *Alpheus heterochelis*. *Canadian Journal of Zoology, 76*, 1080-1090.

Reichmann, V., Irion, U., Wilson, R., Grosskortenhaus, R., & Leptin, M. (1997). Control of cell fates and segmentation in the *Drosophila* mesoderm. *Development, 124*, 2915-2922.

Robertson, M. (1990). More to muscle than MyoD. *Nature, 344*, 378-379.

Sabourin, L. A., & Rudnicki, M. A. (2000). The molecular regulation of myogenesis. *Clinical Genetics, 57*, 16-25.

Sandeman, R., & Sandeman, D. (1991). Stages in the development of the fresh-water crayfish *Cherax destructor*. *Roux's Archives of Developmental Biology, 200*, 27-37.

Schafer, B. W., Czerny, T., Bernasconi, M., Genini, M., & Busslinger, M. (1994). Molecular cloning and characterization of a human *PAX-7* cDNA expressed in normal and neoplastic myocytes. *Nucleic Acids Research, 22*(22), 4574-4582.

Schmidt, M., Tanaka, M., & Munsterberg, A. (2000). Expression of ß-catenin in the developing chick myotome is regulated by myogenic signals. *Development, 127*, 4105-4113.

Scholtz, G. (1995). Expression of the *Engrailed* gene reveals nine putative segment-anlagen in the embryonic pleon of the freshwater crayfish *Cherax destructor* (Crustacea, Malacostraca, Decapoda). *Biological Bulletin, 188*, 157-165.

Scholtz, G. (2000). Evolution of the nauplius stage in malacostracan crustaceans. *Journal of Zoological Systematics and Evolutionary Research, 38*, 175-187.

Scholtz, G., & Dohle, W. (1996). Cell lineage and cell fate in crustacean embryos – a comparative approach. *International Journal of Developmental Biology, 40*(1), 211-220.

Scholtz, G., & Kawai, T. (2002). Aspects of embryonic and postembryonic development of the Japanese freshwater crayfish *Cambaroides japonicus* (Crustacea, Decapoda) including a hypothesis on the evolution of maternal care in the Astacida. *Acta Zoologica, 83*, 203-212.

Seale, P., Polesskaya, A., & Rudnicki, M. A. (2003). Adult stem cell specification by wnt signaling in muscle regeneration. *Cell Cycle, 2*(5), 418-419.

Seale, P., Sabourin, L. A., Girgis-Gabardo, A., Mansouri, A., Gruss, P., & Rudnicki, M. A. (2000). Pax7 is required for the specification of myogenic satellite cells. *Cell, 102*, 777-786.

Seale, P., & Rudnicki, M. A. (2000). A new look at the origin, function, and "stem-cell" status of muscle satellite cells. *Developmental Biology, 218*, 115-124.

Seo, H.-C., Saetre, B. O., Havik, B., Ellingsen, S., & Fjose, A. (1998). The zebrafish *Pax3* and *Pax7* homologues are highly conserved, encode multiple isoforms and show dynamic segment-like expression in the developing brain. *Mechanisms of Development, 70*, 49-63.

Seoighe, C., Johnston, C. R., & Shields, D. C. (2003). Significantly different patterns of amino acid replacement after gene duplication as compared to after speciation. *Molecular Biology and Evolution, 20*(4), 484-490.

Siveter, D. J., Williams, M., & Waloszek, D. (2001). A phosphatocopid crustacean with appendages from the Lower Cambrian. *Science, 293*, 479-481.

Slack, J. M. W. (2003). Regeneration research today. *Developmental Dynamics, 226*, 162-166.

Snider, L. & Tapscott, S. J. (2003). Emerging parallels in the generation and regeneration of skeletal muscle. *Cell, 113*, 811-812.

Spring, J., Yanze, N., Josch, C., Middel, A. M., Winninger, B., & Schmid, V. (2002) Conservation of *Brachyury, Mef2,* and *Snail* in the myogenic lineage of jellyfish: A connection to the mesoderm of bilateria. *Developmental Biology, 244*, 372-384.

Sun, H., Merugu, S., Gu, X., Kang, Y. Y., Dickinson, D. P., Callaerts, P., & Li, W.-H. (2002). Identification of essential amino acid changes in paired domain evolution using a novel combination of evolutionary analysis and in vitro and in vivo studies. *Molecular Biology and Evolution, 19*(9), 1490-1500.

Sun, H., Rodin, A., Zhou, Y., Dickinson, D. P., Harper, D. E., Hewett-Emmett, D., & Li, W.-H. (1997). Evolution of paired domains: Isolation and sequencing of jellyfish and hydra *Pax* genes related to *Pax-5* and *Pax-6*. *Proceedings of the National Academy of Sciences of the United States of America, 94*, 5156-5161.

Tajbakhsh, S. (2003). Stem cells to tissue: molecular, cellular and anatomical heterogeneity in skeletal muscle. *Current Opinion in Genetics & Development, 13*, 413-422.

Tamaki, T., Akatsuka, A., Ando, K., Nakamura, Y., Matsuzawa, H., Hotta, T., Roy, R. R., & Edgerton, V. R. (2002). Identification of myogenic-endothelial progenitor cells in the interstitial spaces of skeletal muscle. *Journal of Cell Biology, 157*(4), 571-577.

Tell, G., Scaloni, A., Pellizzari, L., Formisano, S., Pucillo, C., & Damante, G. (1998). Redox potential controls the structure and DNA binding activity of the paired domain. *Journal of Biological Chemistry, 273*(39), 25062-25072.

Theisen, H., Haerry, T. E., O'Connor, M. B., & Marsh, J. L. (1996). Developmental territories created by mutual antagonism between Wingless and Decapentaplegic. *Development, 122*, 3939-3948.

Turbeville, J. M., Pfeifer, D. M., Field, K. G., & Raff, R. A. (1991). The phylogenetic status of arthropods, as inferred from 18S rRNA sequences. *Molecular Biology and Evolution, 8*(5), 669-686.

Uhrik, B., Rydlova, K., & Zacharova, D. (1989). The role of haemocytes during degeneration and regeneration of crayfish muscle fibres. *Cell and Tissue Research, 255*, 443-449.

Underhill, D. A., & Gros, P. (1997). The paired-domain regulates DNA binding by the homeodomain within the intact Pax-3 protein. *Journal of Biological Chemistry, 272*(22), 14175-14182.

Vogan, K. J., & Gros, P. (1997). The C-terminal subdomain makes an important contribution to the DNA binding activity of the Pax-3 paired domain. *Journal of Biological Chemistry, 272*(45), 28289-28295.

Vogan, K. J., Underhill, D. A., & Gros, P. (1996). An alternative splicing event in the Pax-3 paired domain identifies the linker region as a key determinant of paired domain DNA-binding activity. *Molecular and Cellular Biology, 16*(12), 6677-6686.

West, J. M. (1997). Ultrastructural and contractile activation properties of crustacean muscle fibres over the moult cycle. *Comparative Biochemistry and Physiology, 117B*(3), 333-345.

West, J. M., Humphries, D. C., & Stephenson, D. G. (1995). Characterisation of ultrastructural and contractile activation properties of crustacean (*Cherax destructor*) muscle fibres during claw regeneration and moulting. *Journal of Muscle Research and Cell Motility, 16*, 267-284.

Wiechaus, E., & Nusslein-Volhard, C. (1986). Looking at embryos. In D. B. Roberts (Ed.), *Drosophila: a practical approach* (pp. 199-226). Oxford: IRL Press.

Williams, T. A., & Nagy, L. M. (2001). Developmental modularity and the evolutionary diversification of arthropod limbs. *Journal of Experimental Zoology (Molecular Development and Evolution), 291*, 241-257.

Wilson, D. S., Guenther, B., Desplan, C., & Kuriyan, J. (1995). High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. *Cell, 82*, 709-719.

Xu, H. E., Rould, M. A., Xu, W., Epstein, J. A., & Maas, R. L. (1999). Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes & Development, 13*, 1263-1275.

Xu, W., Rould, M. A., Jun, S., Desplan. C., & Pabo, C. O. (1995). Crystal structure of a paired domain-DNA complex at 2.5 Å resolution reveals structural basis for Pax developmental mutations. *Cell, 80*, 639-650.

Xue, L., Li, X., & Noll, M. (2001). Multiple protein functions of paired in *Drosophila* development and their conservation in the Gooseberry and Pax3 homologs. *Development, 128*, 395-405.

Xue, L., & Noll, M. (1996). The functional conservation of proteins in evolutionary alleles and the dominant role of enhancers in evolution. *European Molecular Biology Organisation Journal, 15*, 3722-3731.

Xue, L., & Noll, M. (2002). Dual role of the Pax gene *paired* in accessory gland development of *Drosophila*. *Development, 129*, 339-346.

Zammit, P. S., & Beauchamp, J. R. (2001). The skeletal muscle satellite cell: stem cell or son of stem cell? *Differentiation, 68*, 193-204.

Zhang, J.-M., Chen, L., Krause, M., Fire, A., & Paterson, B. M. (1999). Evolutionary conservation of MyoD function and differential utilization of E proteins. *Developmental Biology, 208*, 465-472.

Zhang, Y., Ungar, A., Fresquez, C., & Holmgren, R. (1994). Ectopic expression of either the *Drosophila gooseberry-distal* or *proximal* gene causes alterations of cell fate in the epidermis and central nervous system. *Development, 120*, 1151-1161.

Ziman, M. R., & Kay, P. H. (1998). Differential expression of four alternate *Pax7* paired box transcripts is influenced by organ- and strain-specific factors in adult mice. *Gene, 217*, 77-81.

Zorzano, A., Kaliman, P., Guma, A., & Palacin, M. (2003). Intracellular signals involved in the effects of insulin-like growth factors and neuregulins on myofibre formation. *Cell Signal, 15*(2), 141-149.

7.0 APPENDIX

Alignment of Cdpax3/7 and representative Pax protein partial PD (86 amino acids) and partial HD (48 amino acids) used for maximum likelihood phylogenetic analysis.

```
            PD→
Pax3        LRVSHGCVSK ILCRYQETGS IRPGAIGGSK PKVTTPDVEK KIEEYKRENP
Pax7        LRVSHGCVSK ILCRYQETGS IRPGAIGGSK PRVATPDVEK KIEEYKRENP
Tupax       LRVSHGCVSK ILNRYQETGS IRPGVIGGSK PRVATPEVEK KIEEYKRDNP
prd         LRVSHGCVSK ILNRYQETGS IRPGVIGGSK PRIATPEIEN RIEEYKRSSP
gsb         LRVSHGCVSK ILNRFQETGS IRPGVIGGSK PRVATPDIES RIEELKQSQP
gsb-n       LRVSHGCVSK ILNRYQETGS IRPGVIGGSK PKVTSPEIET RIDELRKENP
Cdpax       LRVSHGCVSK ILNRYQETGS IRPGVIGGSK PKVTTPDIEK RIYDYKKENP
pby1        LRVSHGCVSK ILNRYQETGS IRPGVIGGSK PRVATPEVEA RIDDYKKANP
pby2        LRVSHGCVSK ILNRYQETGS IRPGVIGGSK PRVATPEIEA RIEEYKKANP
PaxD        LRVSHGCVSK ILCRYQETGT VDPGIVGLNR PRDVTPEIEN KIDQFRKENS
amphipax37  LRVSHGCVSK ILCRYQETGS IKPGAIGGSK PRVATPEVEK KIEDYKRDNP
PaxB        LRVSHGCVSK ILCRFYETGS IKPGVIGGSK PKVATGPVVN KIAEYKRNNP
Pax5        LRVSHGCVSK ILGRYYETGS IKPGVIGGSK PKVATPKVVE KIAEYKRQNP
Pax2        LRVSHGCVSK ILGRYYETGS IKPGVIGGSK PKVATPKVVD KIAEYKRQNP
Pax8        LRVSHGCVSK ILGRYYETGS IRPGVIGGSK PKVATPKVVE KIGDYKRQNP
amphipax2   LRVSHGCVSK ILRRYYETGS IKPGVIGGSK PKVATPKVVE KIAEYKRQNP
HrPax-258   LRVSHGCVSK ILARYYETGS IKPGVIGGSK PKVATPRVVE KICEYKRQNP
sparkling   LRVSHGCVSK ILSRYYETGS FKAGVIGGSK PKVATPPVVD AIANYKRENP
Pax9        LRVSHGCVSK ILARYNETGS ILPGAIGGSK PRVTTPTVVK HIRTYKQRDP
Pax1        LRVSHGCVSK ILARYNETGS ILPGAIGGSK PRVTTPNVVK HIRDYKQGDP
amphipax1   LRVSHGCVSK ILARYNETGS ILPGAIGGSK PRVTTPEVVK AIKKYKTLDP
PaxA        LLVSHGCVSK ILGRYYETGS VRPGAIGGSK PKVATPRVVS KILAYKEDNP
Poxmeso     LRVSHGCVSK ILARYHETGS ILPGAIGGSK PRVTTPKVVN YIRELKQRDP
Poxneuro    LLVSHGCVSK ILTRFYETGS IRPGSIGGSK TKVATPTVVK KIIRLKEENS
PaxC        LLVSHGCVSK ILGRFYETGS IRPGSIGGSK PKVATPPVVN KIVQYKQQNP
Pax4        LKVSNGCVSK ILGRYYRTGV LEPKCIGGSK PRLATPAVVA RIAQLKDEYP
eyeless     LQVSNGCVSK ILGRYYETGS IRPRAIGGSK PRVATPTVVS KIEQYKRENP
amphipax6   LQVSNGCVSK ILGRYYETGS IRPRAIGGSK PRVATPEVVA KIAQFKRECP
Pax6        ENVSNGCVSK ILGRYYETGS IRPRAIGGSK PRVATPEVVS KIAQYKRECP

            GMFSWEIRDK LLKDAVCDRN TVPSVSSISR ILRSKF
            GMFSWEIRDR LLKDGHCDRS TVPSVSSISR VLRIKF
            GIFAWEIRDR LCKEGICDKN TAPSISSITR LLRTSK
            GMFSWEIREK LIREGVCDRS TAPSVSAISR LVRGRD
            GIFSWEIRAK LIEAGVCDKQ NAPSVSSISR LLRGSS
            SIFSWEIREK LIKEGFAD-- -PPSTSSISR LLRGSD
            GIFSWEIRDD TIKEGVVDKA SAPSVSSISR ILRGGK
            GIFSWEIRDR LIKDGVCDKN SVPSVSSISR LLRGGR
            GIFSWEIRDR LIKEGVCDAN TAPSASAISR LLRGAR
            GIFSWEVRDR LLRENICSKS TVPSLGAISQ ILKSKI
            GMFSWEIRDR LLKDGMCDRS TVPSVSSISR ILRGKG
            TMFAWEIRDR LLSEGVCSTD NVPSVSSINR IVRNRI
            TMFAWEIRDR LLAERVCDND TVPSVSSINR IIRTKV
            TMFAWEIRAQ LLREGICDND TVPSVSSINR IIRTKV
            TMFAWEIRDR LLAEGVCDND TVPSVSSINR IIRTKV
            TMFAWEIRDR LLAEGICDND TVPSVSSINR IVRNKA
            TMFAWEIRDR LLVECICDTE NVPSVSSINR IVRDKA
            TMFAWEIRDR LLAEAICSQD NVPSVSSINR IVRNKA
            GIFAWEIRDR LLADGVCDKY NVPSVSSISR ILRNKI
            GIFAWEIRDR LLADGVCDKY NVPSVSSISR ILRNKI
            GIFAWEIRDR LLAEGVCDKY NVPSVSSISR ILRNKI
            CIFAWEIRNN LLSDGVCDKS NVPSVSSINR ILRNAA
```

```
GIFAWEIRDR LLSEGICDKT NVPSVSSISR ILRNKL
GMFAWEIREQ LQQQRVCDPS SVPSISSINR ILRNSG
TIFAWEIRDR LVEEGVCDRE NTPSVSSINR ILRNKA
ALFAWEIQHQ LCTEGLCTQD KAPSVSSINR VLRALQ
TIFAWEIRER LISEGVCTNA TAPSVSSINR ILRNRA
SIFAWEIRDR LLSEGICTNE NIPSVSSINR VLRNLA
SIFAWEIRDR LLSEGVCTND NIPSVSSINR VLRNLA

HD→
QRRSRTTFTA EQLEELERAF ERTHYPDIYT REELAQRAKL TEARVQVW
QRRSRTTFTA EQLEELEKAF ERTHYPDIYT REELAQRTKL TEARVQVW
QRRSRTTFTA EQLEELEKAF ERSQYPRRI- ---------- --------
QRRCRTTFSA SQLDELERAF ERTQYPDIYT REELAQRTNL TEARIQVW
QRRSRTTFSN DQIDALERIF ARTQYPDVYT REELAQSTGL TEARVQVW
QRRSRTTFTA EQLEALERAF SRTQYPDVYT REELAQTTAL TEARIQVW
QRRSRTTFTA EQLEVLERSF EKTQYPDVYT REELAQKAKL TEARIQVW
QRRSRTTFSG DQLETLERAF QRTQYPDVYT REELAQKTKL TEARVQVW
QRRSRTTFSG DQLETLERAF QRTQYPDVYT REELARRTGL SEARIQVW
QRRSRTKFTH AQLNALEKAF QKTQYPDVYT REELAHRLSL TEARVQVW
QRRSRTTFTP EQLEELEKAF ERTHYPDIYT RELAQRTKLT TEARVQVW
VRRQRTTFSG EQIEQLEKTF EKTHYPDVFT REKLAQDVDL SEARIQVW
------LFTQ QQLEVLDRVF ERQHYSDIFT TTE------- --------
------TFTQ QQLEALDRVF ERPSYPDVFQ ASE------- --------
------TFSQ HHLEALECPF ERQHYPEAYA SPS------- --------
KRS---TFTP DQLEALEQAF NRGHYPTDPF NRDNMS---- --------
---------- ---------- ---------- ---------- - --------
---------- ---------- ---------- ---------- --------
---------- ---------- ---------- ---------- --------
---------- ---------- ---------- ---------- --------
---------- ---------- ---------- ---------- --------
---------- ---------- ---------- ---------- --------
---------- ---------- ---------- ---------- --------
---------- ---------- ---------- ---------- --------
IRRNRTTFSP EQLEMLEKEF EKSHYPDVAT REELASKIDM SEARVQVW
SHRNRTIFSP GQAEALEKEF QRGQYPDSVA RGKLAAATSL PEDTVRVW
FRRNRTTFSP EQLEELEKEF DKSHYPCVST RERLSSRTSL SEARVQVW
LQRNRTSFTQ EQIEALEKEF ERTHYPDVFA RERLAAKIDL PEARIQVW
LQRNRTSFTQ EQIEALEKEF ERTHYPDVFA RERLAAKIDL PEARIQVW
```