

INTER-SPECIES COMPARISON OF THE PEA3 PROMOTER

**INTER-SPECIES COMPARISON OF PROMOTER
SEQUENCES OF THE Ets TRANSCRIPTION
FACTOR PEA3**

**By
GREGORY SCOTT KANN, B.Sc.**

**A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree
Master of Science**

McMaster University

© Copyright by Gregory Scott Kann, September 1999

MASTER OF SCIENCE (1999)
(Biochemistry)

McMaster University
Hamilton, Ontario

TITLE: Inter-species Comparison of Promoter Sequences of the Ets Transcription Factor PEA3

AUTHOR: Gregory Scott Kann, B.Sc. (McMaster University)

SUPERVISOR: Professor J.A. Hassell

NUMBER OF PAGES: xii, 232

ABSTRACT

Chicken and pufferfish genomic libraries were screened with the intent of isolating PEA3 orthologues from evolutionarily removed vertebrate species. The chicken PEA3 gene was found to reside within 15 kb of genomic sequence, and approximately 2 kb of promoter sequence has been identified. Although the pufferfish PEA3 genomic sequence has yet to be completed, exons 2, 3, 4, 5, 12 and 13 have been found, and approximately 1 kb of sequence upstream of the putative start codon has been determined. In addition to the genomic sequence that was isolated, 5' RACE using pufferfish heart RNA produced a 334 bp cDNA sequence encompassing exons 2 to 5 of pufferfish PEA3. A pufferfish homologue of the human RNA helicase 1 (HRH1) gene was also found 3' of the PEA3 gene. Given that HRH1 is also found 3' of human PEA3 (E1AF) on chromosome 17q21, this finding would seem to indicate that chromosomal synteny is maintained between the human and pufferfish PEA3 loci. A four-way alignment of the mouse, human, chicken and pufferfish PEA3 promoters revealed that a region spanning from +1 to -260, relative to the transcriptional start site of mouse PEA3, is well conserved across the four promoters. Conserved transcription factor binding sites for SRY, HNF3 β , NFY, AP-1, TCF, AP-2, v-myb, δ EF1, and c-Ets-1 were found in three, and in some cases four of the promoters. An additional outcome of the pufferfish genomic library screen was the isolation of a pufferfish orthologue of the Ets transcription factor ERM. The relevance of these findings to the issue of transcriptional regulation of PEA3 expression is discussed.

ACKNOWLEDGMENTS

I would like to thank my graduate supervisor John A. Hassell for giving me the opportunity to pursue my scientific research interests. Throughout my studies I was encouraged to think critically about all aspects of research and I believe that this has prepared me to further pursue my future goals. I would also like to thank the members of the Hassell lab, both past and present, for all of their support.

CONTRIBUTIONS

I would like to thank the following individuals for their contributions to my thesis research. DNA sequencing and synthesis of oligonucleotide primers was carried out by Brian Allore and Dinsdale Gooden in the Central Facility at the Institute for Molecular Biology and Biotechnology (MOBIX), McMaster University. Isolation of pufferfish PEA3 genomic cosmids from a gridded genomic pufferfish library, and generation of a 5' partial cDNA from pufferfish PEA3 by 5' RACE, was carried out by Jonathon Gilley in the laboratory of Dr. M. Fried at the Imperial Cancer Research Fund, London, England. Additional DNA sequencing was carried out by Ji-Hou Xin, Jane Barrett and Bonnie Bojovic.

TABLE OF CONTENTS

ABSTRACT	iii
ACKNOWLEDGEMENTS	iv
CONTRIBUTIONS	v
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF APPENDICES	xii
INTRODUCTION	1
1. The Ets Family of Transcription Factors	1
2. The PEA3 Group	5
3. PEA3 Binding Sites are Found in a Variety of Promoters	9
4. Ets Transcription Factors and Disease	10
5. Cross Species Comparison of PEA3 Promoters	13
6. The Japanese Pufferfish <i>Fugu rubripes</i> as a Mechanism for Promoter Study	14
MATERIALS	20
METHODS	21
1. Isolation and Characterization of Chicken PEA3 Cosmids	21
a. Library Screening	21
b. PCR Generation of a Chicken PEA3 Specific Genomic Probe	24
c. Large Scale DNA Preparation	25
d. Restriction Digestion of Putative Positive Cosmids	26
e. Southern Blot Analysis of Cosmid DNA	27
f. DNA Sequencing Using Chicken PEA3 Specific Primers	27
g. Subcloning of 5' Genomic Chicken PEA3 Restriction Digest Fragments	28
2. Isolation and Characterization of Pufferfish ERM Bacteriophage λ Clones	30
a. Library Screening	30
b. λ Bacteriophage DNA Purification	32
c. Restriction Digestion of Putative Positive λ Bacteriophage Inserts	33
d. Southern Blot Analysis of λ Bacteriophage DNA	33
e. PCR Generation of a Pufferfish ERM Specific Genomic Probe	34
f. Subcloning of DNA Fragments Containing Pufferfish ERM	34
g. DNA Sequencing of Pufferfish ERM	36
3. Reverse Transcriptase Polymerase Chain Reaction Isolation of Zebrafish PEA3	36

4. Isolation and Characterization of Pufferfish PEA3 Cosmids	38
a. Library Screening	38
b. PCR Generation of a 5' Zebrafish PEA3 cDNA Probe	39
c. PCR Generation of a Pufferfish PEA3 Specific Genomic Probe	40
d. 5' RACE Using a Pufferfish PEA3 Exon 5 Primer	40
e. Large Scale DNA Preparation	40
f. Restriction Digestion of Putative Positive Cosmids	41
g. Southern Blot Analysis of Cosmid DNA	41
h. Subcloning of Pufferfish PEA3 Restriction Digest Fragments	42
i. DNA Sequencing of Pufferfish PEA3	42
5. Computer Assisted Analysis of Genomic Sequence	43
a. Sequence Editing	43
b. Multiple Sequence Alignment	43
c. BLAST (Basic Local Alignment Search Tool) Analysis of Sequence Data	43
d. Identification of Consensus Transcription Factor Binding Sites Within Promoter Sequences	44
RESULTS	45
1. Isolation and Characterization of Chicken PEA3 Cosmids	45
a. Library Screening and PCR Generation of a Chicken PEA3 Specific Genomic Probe	45
b. Restriction Digestion of Putative Positive Cosmids	48
c. Southern Blot Analysis of Cosmid DNA	53
d. DNA Sequencing Using Chicken PEA3 Specific Primers	58
e. Subcloning of 5' Genomic Chicken PEA3 Restriction Digest Fragments	59
2. Reverse Transcriptase Polymerase Chain Reaction Isolation of Zebrafish PEA3	62
3. Isolation and Characterization of Pufferfish PEA3 Cosmids	67
a. Library Screening and PCR Generation of 5' Zebrafish PEA3 cDNA and Genomic Pufferfish PEA3 Probes	67
b. 5' RACE Using a Pufferfish PEA3 Exon 5 Primer	72
c. Restriction Digestion of Putative Positive Cosmids	72
d. Southern Blot Analysis of Cosmid DNA	79
e. Subcloning of Pufferfish PEA3 Restriction Digest Fragments	84
f. DNA Sequencing of Pufferfish PEA3	91
4. Four-way Alignment of The Mouse, Human, Chicken and Pufferfish PEA3 Promoters ..	94
a. Sequence Identity Between the Promoters	94
b. Identification of Conserved Consensus Transcription Factor Binding Sites Within the Promoters	95
c. Common Features Within the PEA3 Promoters	96

5. Isolation and Characterization of Pufferfish ERM Bacteriophage λ Clones	113
a. Library Screening and PCR Generation of a Pufferfish ERM Specific Genomic Probe	113
b. Restriction Digestion of Putative Positive λ Bacteriophage Inserts	113
c. Southern Blot Analysis of λ Bacteriophage DNA	121
d. Subcloning of DNA Fragments Containing Pufferfish ERM	126
e. DNA Sequencing of Pufferfish ERM	126
f. Computer Assisted Identification of Pufferfish ERM	129
6. Two-way Alignment of the Human and Pufferfish ERM Promoters	130
a. Sequence Identity Between the Promoters	130
b. Identification of Conserved Consensus Transcription Factor Binding Sites Within the Promoters	137
DISCUSSION	139
1. Isolation and Sequencing of the Chicken PEA3 Locus	139
a. Size of Genomic Chicken PEA3	139
b. Intron-Exon Organization of Chicken PEA3	139
c. Conservation of Putative Phosphorylation Sites in Chicken PEA3	140
2. Isolation and Sequencing of the Pufferfish PEA3 Locus	142
a. Partial Size and Intron-Exon Organization of Genomic Pufferfish PEA3	142
b. Conservation of Chromosomal Synteny at the PEA3 Locus	144
c. Identification of a 5' Partial cDNA of Pufferfish PEA3	145
3. Alignment of PEA3 Orthologue Promoters	150
a. Observed Sequence Identity	150
b. Conserved Transcription Factor Binding Sites	151
c. Common Features Within the PEA3 Promoters	159
d. Missing Features Within the Pufferfish PEA3 Promoter	161
4. Isolation and Sequencing of the Pufferfish ERM Locus	162
a. Size of Genomic Pufferfish ERM	162
b. Intron-Exon Organization of Pufferfish ERM	163
5. Two-way Alignment of the Human and Pufferfish ERM Promoters	163
a. Sequence Identity Between the Promoters	163
b. Identification of Conserved Consensus Transcription Factor Binding Sites Within the Promoters	164
CONCLUSION	166
REFERENCES	167

LIST OF TABLES

Table 1	Sub-groups of the Ets Transcription Factor Family	3
Table 2	Current Predictions of Approximate Gene Number and Genome Size in Organisms of Varying Evolutionary Lineage	15
Table 3	Percent Similarity and Percent Divergence Between the Amino Terminal End of Pufferfish PEA3 and Other PEA3 Group Proteins	148

LIST OF FIGURES

Figure 1	Evolution of the PEA3 Group of Transcription Factors	6
Figure 2	Timeline of Vertebrate Evolution	17
Figure 3	PCR Generation of a Genomic Chicken PEA3 Specific Probe	46
Figure 4	Restriction Digestion of Putative Positive Chicken Cosmids from the Initial Library Screen	49
Figure 5	Restriction Digestion of Putative Positive Chicken Cosmids from the Second Library Screen	51
Figure 6	Southern Blot Analysis of Putative Positive Chicken Cosmids from the Initial Library Screen	55
Figure 7	Southern Blot Analysis of Putative Positive Chicken Cosmids from the Second Library Screen	56
Figure 8	A Summary of the Subcloned Restriction Digestion Fragments from the Chicken PEA3 Cosmids	60
Figure 9	Zebrafish PEA3 cDNA Sequence	63
Figure 10	Zebrafish PEA3 cDNA RT-PCR	65
Figure 11	PCR Generation of a 5' Zebrafish PEA3 cDNA Probe	68
Figure 12	PCR Generation of a Genomic Pufferfish PEA3 Probe	70
Figure 13	Restriction Digestion of a Putative Positive Pufferfish Cosmid from the Initial Library Screen	73
Figure 14	Restriction Digestion of a Putative Positive Pufferfish Cosmid from the Second Library Screen	75
Figure 15	Restriction Digestion of Two Further Putative Positive Pufferfish Cosmids from the Second Library Screen	77
Figure 16	Restriction Digestion of Putative Positive Pufferfish Cosmids from the Third Library Screen	80
Figure 17	Southern Blot Analysis of the Putative Positive Pufferfish Cosmid pPEA3-1	82
Figure 18	Southern Blot Analysis of the Putative Pufferfish Cosmid pPEA3-1A	85
Figure 19	Southern Blot Analysis of the Putative Pufferfish Cosmids pPEA3-1B and pPEA3-1C	87

Figure 20	Southern Blot Analysis of the Putative Positive Pufferfish Cosmids from the Third Library Screen	89
Figure 21	A Summary of Subcloned Restriction Fragments from the Pufferfish PEA3 Cosmids	92
Figure 22	Four-way Alignment of the Mouse, Human, Chicken and Pufferfish PEA3 Promoters	97
Figure 23	PCR Generation of a Pufferfish ERM Specific Genomic Probe	114
Figure 24	Restriction Digestion of Putative Positive Pufferfish λ Bacteriophage Clones	116
Figure 25	Restriction Digestion and Southern Blot Analysis of the Putative Pufferfish ERM λ Clone pERM-A	119
Figure 26	Southern Blot Analysis of Genomic Pufferfish λ Bacteriophage DNA from the Initial Library Screen	122
Figure 27	Southern Blot Analysis of Genomic Pufferfish λ Bacteriophage DNA from the Second Library Screen	124
Figure 28	A Summary of Subcloned Restriction Fragments Containing Pufferfish ERM	127
Figure 29	Two-way Alignment of the Human and Pufferfish ERM Promoters 5' of the Sequence Break	131
Figure 30	Two-way Alignment of the Human and Pufferfish ERM Promoters 3' of the Sequence Break	135
Figure 31	Alignment of the Deduced Partial Amino Acid Sequence of Pufferfish PEA3 with Other PEA3 Group Proteins	146

LIST OF APPENDICES

Appendix A	Chicken PEA3 Genomic Sequence	174
Appendix B	The Pufferfish PEA3 5' RACE Product	189
Appendix C	Pufferfish PEA3 Genomic Sequence at the 5' End of the Gene	191
Appendix D	Pufferfish PEA3 Genomic Sequence Within the Gene	195
Appendix E	Pufferfish PEA3 Genomic Sequence at the 3' End of the Gene	200
Appendix F	Pufferfish ERM Genomic Sequence	204
Appendix G	Two-way Alignment of the Mouse and Pufferfish PEA3 Promoters	211
Appendix H	Two-way Alignment of the Human and Pufferfish PEA3 Promoters	219
Appendix I	Two-way Alignment of the Chicken and Pufferfish PEA3 Promoters	227

INTRODUCTION

The Ets Family of Transcription Factors

The original v-ets (E26 transformation specific) oncogene was first discovered as part of a tripartite fusion protein expressed by the E26 avian erythroblastosis virus (LePrince et al., 1983; Nunn et al., 1983). This replication deficient virus expresses a 5.7 kb RNA that encodes a 135 kDa gag-myb-ets hybrid protein where the v-ets sequence is fused to the viral gag gene, as well as a virally transduced form of myb (Seth et al., 1992). Whereas the closely related avian myeloblastosis virus (AMV) harbors the v-myb oncogene and, consequently, is only capable of inducing myeloblastosis, the E26 virus is able to induce both erythroblastosis and low levels of myeloblastosis in chickens. It has been demonstrated that the Ets portion of the E26 expressed fusion protein is necessary for induction of erythroblastosis (Nunn & Hunter, 1989) and that it also plays a role in the transformation of myeloid cells (Golay et al., 1988). Since the original discovery of the v-ets oncogene, a large family of cellular counterparts have emerged. This emergent Ets gene family has been shown to play an ever expanding role in a wide range of biological processes including embryonic development, response to extracellular signaling and cell transformation (Scott et al., 1994; Bories et al., 1995; Muthusamy et al., 1995).

To date, the still growing Ets family of transcription factors is comprised of over 30 members encompassing genes that have been isolated from a variety of organisms ranging from *Drosophila* (Macleod et al., 1992; Wasyluk et al., 1993) and *C. elegans* (Beitel et al., 1995) to humans (Laudet et al., 1999). The Ets genes that have been identified in vertebrates include c-Ets-1 and c-Ets-2 (Watson et al., 1988), Fli-1 (Ben-David et al., 1991), ERG (Reddy et al., 1987; Rao et al., 1987), FEV (Peter et al., 1997), ER71 (Brown & McKnight, 1992), ER81/ETV1 (Brown & McKnight, 1992), ERM/ETV5 (Monte et al., 1994), PEA3/E1AF/ETV4 (Xin et al., 1992; Higashino et al., 1993), PU.1/Spi-1 (Klemsz et al., 1990; Goebel et al., 1990), Spi-B (Ray et al., 1992), SAP-1 (Dalton et al., 1992), NET/ERP/SAP-2 (Lopez et al.,

1994), Elk-1 (Rao et al., 1989), ERF (Liu et al., 1997), Elf-1 (Thompson et al., 1992), MEF (Mao et al., 1999), Nerf (Oettgen et al., 1996), TEL (Golub et al., 1994), ESX (Chang et al., 1997), and GABP α /E4TF1 (LaMarco et al., 1991). Within the Ets gene family further sub-classification may be carried out on the basis of sequence similarity within the Ets domain, the position of the Ets domain in the protein, and additional similarity throughout the remainder of the amino acid sequence (Wasylyk et al., 1993). Using this criteria, the Ets family has been sub-divided into 13 distinct groups (Laudet et al., 1999) (Table 1).

Common to all of the protein products of the Ets gene family is a highly conserved DNA binding domain called the Ets domain. This signature element, which is contained in all Ets transcription factors, spans 85 amino acids and demonstrates strong conservation throughout the family (Karim et al., 1990). Sequence identity between the most divergent Ets domains (from human Ets-1 and PU.1) reaches 35% (Wasylyk et al., 1993). Inspection of the primary amino acid sequence within Ets domains from a wide variety of Ets factors has revealed the presence of three conserved tryptophan residues that are common to all members of the Ets family with the exception of PU.1 (Macleod et al., 1992).

The Ets domain may be divided into two distinct regions: an amino-terminal portion which contains an α -helical region common to all members of the Ets gene family; and a carboxy-terminal portion which contains a basic domain that also exhibits striking identity throughout the family (Laudet et al., 1993). Comparisons of the configuration of the conserved positions within the Ets domain to other well characterized DNA binding motifs such as zinc fingers, homeodomains, or leucine zippers have revealed no apparent similarity (Johnson & McKnight, 1989; Mitchell & Tjian, 1989). This finding lends support to the conclusion that the Ets domain encodes a new structural motif for DNA binding (Karim et al., 1990). Nuclear magnetic resonance (NMR) studies of the Ets domain from the prototypic Ets transcription factor Ets-1 have shown that the 85 amino acid Ets domain folds into a winged helix-turn-helix motif. This structural motif consists of three α -helices folded into a helix-turn-helix configuration, packed onto a four-stranded anti-parallel β -sheet (Donaldson et al., 1996). Further studies have indicated that the protein-DNA

Table 1 Sub-groups of the Ets Transcription Factor Family

The 13 groups of the Ets transcription factor family are shown. The name of the group is given in the first column, the factors that comprise a group of transcription factors are given in the second column and the organisms where the specific genes have been isolated from are given in the third column.

Group	Gene	Other Names	Organism
PEA3	PEA3	E1AF/ETV4	Mouse, Human, Zebrafish, Chicken (genomic), Pufferfish (partial cDNA)
	ER81	ETV1	Mouse, Human, Xenopus
	ERM	ETV5	Mouse, Human, Chicken, Pufferfish (genomic)
ELF	ELF-1		Mouse, Human
	NERF		Human
	MEF		Human
	E74		<i>Drosophila</i>
ELG	GABP α	E4TF-1	Mouse, Human
	ELG		<i>Drosophila</i>
ELK	SAP-1		Mouse, Human
	NET	ELK-3/ERP/SAP-2	Mouse, Human
	ELK-1		Mouse, Human
	Lin-1		<i>C. elegans</i>
ERF	ERF		Mouse, Human
	PE-1	ETV3/PEP-1	Human
ERG	Fli-1	ERG-B	Mouse, Human, <i>Xenopus</i>
	ERG		Mouse, Human, Sea Urchin
	ETS-6		<i>Drosophila</i>
ETS	ETS-1	•	Mouse, human, Chicken, <i>Xenopus</i> , Rat
	ETS-2		Mouse, Human, Chicken, <i>Xenopus</i> , Sea Urchin
	PNT		<i>Drosophila</i>
SPI	PU.1	SPI-1	Mouse, Human
	SPI-B		Human
YAN	TEL	ETV6	Human
	YAN		<i>Drosophila</i>
ER71	ER71		Mouse, Human
ESX	ESX	ELF3/ESE-1/ERT/JEN	Mouse, Human
FEV	FEV		Human
ETS-4	ETS-4		<i>Drosophila</i>

interaction encompasses a fairly large stretch of 20 base pairs and that the recognition of the core binding sequence occurs in the major groove of the double helix (Donaldson et al., 1996).

A comparison of the DNA sequences bound by Ets factors has revealed that the Ets DNA binding domain binds to a centrally located invariant purine-rich core motif (5' C/A GGA A/T 3') in the middle of a 10 bp recognition element (Wasylyk et al., 1993). The neighboring sequences to this purine-rich core are highly variable amongst different Ets binding sites and it has been hypothesized that this variability confers binding site specificity to Ets proteins. Consistent with this hypothesis, it has been determined that the consensus binding sequence of Ets-1 is 5' A/GCCGGAA/TGT/C 3' (Nye et al., 1992; Woods et al., 1992).

The PEA3 Group

PEA3 (polyomavirus enhancer activator 3), also called E1AF or ETV4, is the founding member of the PEA3 group of Ets transcription factors. Other members of the PEA3 group include ER81 (also called ETV1) (Brown & McKnight, 1992) and ERM (also called ETV5) (Monte et al., 1994). PEA3 was first identified as a factor capable of binding to the PEA3 motif (5' AGGAAG 3'), which was initially found in the polyomavirus enhancer (Martin et al., 1988). This Ets transcription factor was identified in mouse 3T6 nuclear extracts and the cDNA was originally cloned from a mouse FM3A cell library (Xin et al., 1992).

PEA3, ER81 and ERM share a 95% amino acid sequence identity within their respective Ets domains. Further, in the amino-terminal end of the proteins, there is an 85% identical acidic-domain, which is believed to play a role in the transactivation potential of PEA3 (Bojovic & Hassell, private communication). Overall, the three members of the PEA3 group are 50% identical throughout the full length of their protein sequence (Monte et al., 1996). Inspection of the sequence identity within the PEA3 group has revealed that ERM and ER81 are more closely related to each other than they are to PEA3. This finding suggests that a common ancestral gene has undergone two successive duplication events in order to produce the PEA3 group (de Launoit et al., 1997) (Figure 1).

Figure 1 Evolution of the PEA3 Group of Transcription Factors

The two successive duplication events leading to the evolution of the PEA3 group are indicated as red arrows.

It has been determined that the human chromosomal location of PEA3 is at position 17q21 (Isobe et al., 1995; Barrett, 1997). ERM is situated at position 3q27-q29 (Monte et al., 1996), and ER81 at position 7q21. The position of PEA3 is particularly interesting as it is situated in the vicinity of the inherited breast cancer gene BRCA1 (Brody et al., 1995; Osborne-Lawrence et al., 1995). Other genes that lie adjacent to PEA3 include human RNA helicase 1 (directly 3') and MOX1 (lying 5') (Brody et al., 1995). The DNA consensus binding site has been determined for PEA3 and a number of other Ets transcription factors. Similar to Ets-1 (Nye et al., 1992; Woods et al., 1992) the binding site for PEA3 was determined to contain the invariant purine-rich core 5' GGAA/T 3' centered within an approximately 10 bp recognition element. The binding site for PEA3 has been determined to be 5' T/C T/G C C GGA A/T G/C C G 3' (S. Bowman and J.A. Hassell, private communication).

In mouse, the expression pattern of PEA3 has been shown to be highly restricted. As a result, PEA3 mRNA can only be readily detected in the epididymis, the brain and, to a lesser extent, in the mammary gland (Xin et al., 1992). It has been shown however that ERM mRNA is expressed in a much more ubiquitous fashion. For instance, in human tissues, ERM mRNA is expressed very highly in brain and placenta; highly in the lung and pancreas; moderately in heart and skeletal muscle; weakly in kidney; and very weakly in liver (Monte et al., 1994). The expression pattern of ERM has also been examined in mouse tissues where it demonstrates an equally wide tissue distribution in that it is highly expressed in the mouse brain, and moderately expressed in the spleen, testis, and kidney (Monte et al., 1994). The expression pattern of ER81 is similar to that of ERM in that it too shows a wide, somewhat unrestricted, expression pattern. In human tissues for example, ER81 is expressed very highly in brain; highly in testis, lung and heart; moderately in spleen, small intestine, pancreas and colon; weakly in liver, prostate and thymus; very weakly in skeletal muscle, kidney and ovary; and not at all in placenta and blood leukocytes (Monte et al., 1995). ER81 mRNA has also been studied in mouse tissues where it has been detected at high levels in brain and kidney and at moderate levels in testis (Monte et al., 1994).

In addition to adult human and mouse tissues, the expression patterns of the PEA3 group transcription factors have been studied during mouse embryonic development. Specifically, *in situ*

hybridization analysis of the expression patterns of PEA3, ERM and ER81 has revealed a considerable number of sites of co-expression within the developing embryo (Chotteau-Lelievre et al., 1997). Further studies have shown that while PEA3 and ERM exhibit the most striking level of expressional overlap, each factor does possess its own unique sites of expression (Laing & Hassell, private communication). This finding suggests a potential for the existence of similar transcription factor binding sites in the respective promoters, thus governing the overlapping expression. Northern analysis has shown that ERM and ER81 expression continues after the birth of the embryo in numerous organs, while PEA3 expression drops off and is restricted to a relatively small number of organs (Chotteau-Lelievre et al., 1997). Taken together, these results seem to indicate that the PEA3 group plays an important role in mouse embryonic organogenesis.

PEA3 Binding Sites are found in a Variety of Promoters

A number of candidate PEA3 target genes that contain *bonafide* PEA3 recognition sites within their promoters have been identified. To date, this growing group of target genes is comprised almost exclusively of matrix metalloproteinases (MMPs), which are required for degradation of the extracellular matrix. These proteins provide a mechanism through which PEA3 is able to influence metastasis of tumor cells. MMP promoters that contain PEA3 binding sites include gelatinase B (MMP-9), collagenase-3 (MMP-13), stromelysin-1 (MMP-3), stromelysin-2 (MMP-10), stromelysin-3 (MMP-11), matrilysin (MMP-7), and interstitial collagenase (MMP-1) (Matrisian et al., 1994). Despite the presence of PEA3 binding sites in each of the aforementioned MMP promoters, PEA3 has been only found to upregulate transcription from the MMP-1, MMP-9 and MMP-11 promoters (Higashino et al., 1995). PEA3 has also been shown to confer an invasive phenotype on a human breast cancer cell line, MCF-7, that is initially non-metastatic prior to transfection (Kaya et al., 1996). In addition to MMP promoters, PEA3 binding sites have also been found in the urokinase-type plasminogen activator promoter (Rorth et al., 1990). PEA3 has

also been shown to transactivate the Vimentin promoter in mammary epithelial and tumor cells (Chen et al., 1996), as well as the Her-2/*neu* receptor tyrosine kinase promoter (Benz et al., 1996). The closely related PEA3 group member ERM has also been shown to bind with, and transactivate, the intercellular adhesion molecule-1 (ICAM-1) promoter (de Launoit et al., 1998).

Ets Transcription Factors and Disease

Since the original discovery of the v-ets oncogene and its ability to induce erythroleukemia in chickens (LePrince et al., 1983), a wide range of Ets factors have been implicated in the onset of disease in both humans and mice. In the case of v-ets induced oncogenesis, the E26 retrovirus induces erythroblastosis upon expression of a gag-myb-ets fusion protein (LePrince et al., 1983; Nunn et al., 1983). Other cases of retroviral activation of Ets genes are also exhibited in higher vertebrates. Most notably, erythroleukemia has been found to arise in mice due to retroviral insertion of either the Friend murine leukemia virus (F-MuLV), or the Spleen focus forming virus (SFFV), near the locus of the Ets genes *Fli-1* and *Spi-1/PU.1* respectively (Ben-David et al., 1991; Dittmer & Nordheim, 1998). These instances of tumor onset involving Ets genes suggest that deregulation of Ets gene expression may serve as a critical factor during tumorigenesis.

Ets transcription factors also play an important role in the onset of tumorigenesis in that they serve as downstream targets of signaling cascades. In particular, enhanced expression of PEA3 has been shown to facilitate tumor progression and metastasis due to overexpression of the rat *neu* protooncogene in transgenic mice. The Her-2/*neu* protooncogene (also called c-erbB2, and ERBB2) encodes a 185 kDa transmembrane protein that is a member of the epidermal growth factor (EGF) receptor family (Bargmann et al., 1986). Amplification and overexpression of the human homologue of *neu* has been clinically observed in a large portion (30%) of primary breast cancers and it has been determined that this state generally indicates a poor prognosis in affected patients (Slamon et al., 1987). Transgenic mice bearing a

constitutively active *neu* transgene have been shown to rapidly develop mammary tumors that involve the entire mammary epithelium. This result suggests that activated *neu* alone is sufficient to induce transformation of mammary epithelial cells (Muller et al., 1988). In contrast, transgenic mice carrying the unactivated form of *neu* have been shown to develop tumors in a much slower fashion. Once primary tumors have developed however, the tumors were found to be capable of metastasizing to the lung: an observation that was also made in the activated *neu* transgenic mice. This finding indicates that an additional genetic event was required to initiate transformation of the mammary epithelium (Guy et al., 1992). The expression of PEA3 has also been studied in transgenic mice bearing the unactivated form of *neu*. In this case, it was determined that high levels of PEA3 RNA were present in *neu*-induced tumors, with little or no expression being observed in adjacent non-transformed mammary tissue. PEA3 RNA levels were subsequently studied in mammary tumors that had metastasized to the lungs. In this case, it was also determined that PEA3 RNA levels were elevated with no detection of a PEA3 message in the surrounding lung tissue (Trimble et al., 1993). When taken together, these results provide convincing evidence in support of a role for PEA3 in facilitating mammary epithelial cell transformation and metastasis (Trimble et al., 1993). Interestingly, all tissues bearing elevated levels of PEA3 RNA also possessed increased levels of *neu* RNA. This suggests that PEA3 is capable of upregulating endogenous *neu* mRNA levels. Further studies have also shown that PEA3 is capable of transactivating its own promoter, thereby providing a mechanism by which an initial signaling cascade can lead to overexpression of both *neu* and PEA3. Given this overexpression, PEA3 is capable of upregulating additional target genes, such as the matrix metalloproteinases, and it is in this fashion that PEA3 functions to facilitate the onset of tumorigenesis and metastasis (Benz et al., 1996).

Chromosomal translocation that results in the production of chimeric fusion proteins has been implicated in human tumorigenesis. Fusion proteins involving Ets transcription factors have been found to occur in various forms of leukemia, including B-type childhood acute lymphoblastic leukemia (ALL)

(Dittmer & Nordheim, 1998), and Ewing tumors (Delattre et al., 1992; Zucman et al., 1993). Ewing tumors arise from a frequently occurring translocation involving band q12 of human chromosome 22. This locus encodes the EWS (Ewing Sarcoma) protein that contains an RNA binding domain (Zucman et al., 1993).

By far the most common chromosomal translocation involving EWS and an Ets gene is the translocation of band 22q12 to chromosome 11, resulting in the fusion of the EWS gene with the Ets transcription factor Fli-1 gene (May et al., 1993; Braun et al., 1995; Zucman et al., 1993). The resultant hybrid EWS-Fli-1 protein is known to act as a strong transcriptional activator and has been shown to facilitate transformation. In particular, EWS-Fli-1 has been shown to specifically interact via the amino terminus of EWS with the seventh largest subunit (hsRPB7) of human RNA polymerase II. In this fashion this chimeric protein is able to directly affect gene transcription (Petermann et al., 1998).

A second, less commonly observed EWS fusion has been found to occur with the closely related Ets gene, ERG, which is located on chromosome 21 (Zucman et al., 1993). Two other Ets transcription factors, ER81 and PEA3, have also been found to be involved in the onset of Ewing's sarcoma / PNET (primitive neuroectodermal tumors). ER81/EWS fusions arise from the translocation $t(7;22)(p22;q12)$, which fuses the EWS sequence to a portion of ER81 that codes for the Ets DNA binding domain (Jeon et al., 1995). Although PEA3 has also been identified as a translocation partner of EWS (Urano et al., 1996), this translocation appears to occur far less frequently than the other three EWS-ETS translocation events. However, when it does occur, fusion of the EWS sequence to the Ets domain of PEA3 in Ewing's sarcoma produces tumors of a very malignant phenotype, and this is not surprising since PEA3 is known to transactivate the promoters of a number of matrix metalloproteinase genes (Kaya et al., 1996; Higashino et al., 1995).

Another Ets factor that has been found to be involved in chromosomal rearrangement is TEL on chromosome 12, which has been found to fuse with the DNA-binding and transactivation domains of the transcription factor AML1 (located on chromosome 21 at position q22). This translocation, $t(12;21)$, gives rise to childhood acute lymphoblastic leukemia (Golub et al., 1995). TEL has also been found to fuse with

the PDGF receptor β via a t(5;12) translocation, giving rise to chronic myelomonocytic leukemia (Golub et al., 1994). Finally, the Ets transcription factor ERG has also been shown to fuse with the transcription factor TLS on chromosome 16 in some myeloid leukemias (Ichikawa et al., 1994).

Cross Species Comparison of PEA3 Promoters

In order to fully understand which transcription factors play a role in regulating PEA3 expression it is necessary to identify those regions, or specific binding sites, within its promoter that may be relevant. To undertake such a study, it is necessary to examine PEA3 orthologues from a variety of evolutionarily removed species. Such a strategy would allow for the identification of conserved sequences in the PEA3 promoter that could reveal those transcription factor binding sites that are of true importance. Once found, these binding sites could be subjected to further study in the context of the mouse PEA3 promoter, either by promoter mapping studies in transfection experiments or with *in vivo* transgenics, where a specific region of the mouse PEA3 promoter is used to drive expression of a β -galactosidase transgene.

Inspection of the proximal promoter sequences from human and mouse PEA3 indicate a high degree of sequence conservation, however, this high degree of sequence identity makes it difficult to determine precisely which sequence elements are necessary to govern PEA3 expression. Deletion analysis has been somewhat successful in identifying that the core promoter elements lie within the region -156 to +676 (Barrett, 1997). This is surprising to a degree since there are an abundance of potential transcription factor binding sites directly 5' to these sequences. In an attempt to further define the core promoter, the chicken and pufferfish genomic PEA3 homologues were pursued by screening genomic libraries from each organism. Following isolation of these genes the promoter sequences were compared with those from human and mouse.

The Japanese Pufferfish *Fugu rubripes* as a Mechanism for Promoter Study

An ideal system in which to carry out cloning, sequencing, and promoter regulatory element study would be in a genome that is small, but still maintains the same gene complement as in higher vertebrates. The Japanese Pufferfish *Fugu rubripes* provides just such a system. The pufferfish genome differs from many of the conventional organisms of study (such as *E. coli*, yeast, *Drosophila* and *C. elegans*) in that it possesses a gene complement that is extremely similar to that of human and mouse (approximately 70,000 genes) (Miklos et al., 1996) (Table 2). Thus, it is reasonable to assume that all members of the Ets family are present in this genome. The evolutionary distance of the pufferfish from higher vertebrates genome (approximately 400 million years) (Figure 2) can be used to reveal vital sequence conservation within both coding sequences and non-coding elements such as promoters (Brenner et al., 1993). Another advantage to using *Fugu rubripes* as a model genome is that the genome size is reduced by approximately 7 to 8 fold in relation to either human or mouse (pufferfish genome - 400 Mb, human genome - 3300 Mb, mouse genome - 3300 Mb). Given that the gene compliment is the same and that there is a 7 to 8 fold reduction in the size of the pufferfish genome, one can reasonably infer that there is a lower abundance of non-coding DNA. This inference is supported by the finding that the pufferfish genome contains less than 10% repetitive DNA elements and the fact that the G/C content is found to be 44.2%, which is slightly higher than that found in the human genome (40.3%) (Elgar et al., 1996). It is also important to note that, in many cases where pufferfish gene structure has been compared to human or mouse homologues, intron-exon organization was conserved. Sequencing of the pufferfish genome has also shown a high degree of synteny conservation between the pufferfish genome and higher mammals, particularly humans (Elgar et al., 1996).

There are a number of cases in which the pufferfish genome has been used to detect conserved regulatory elements are beginning to emerge. For example, mouse and pufferfish sequence comparisons have been performed for the *Hoxb-4* gene (Aparicio et al., 1995). In this case homologous sequences,

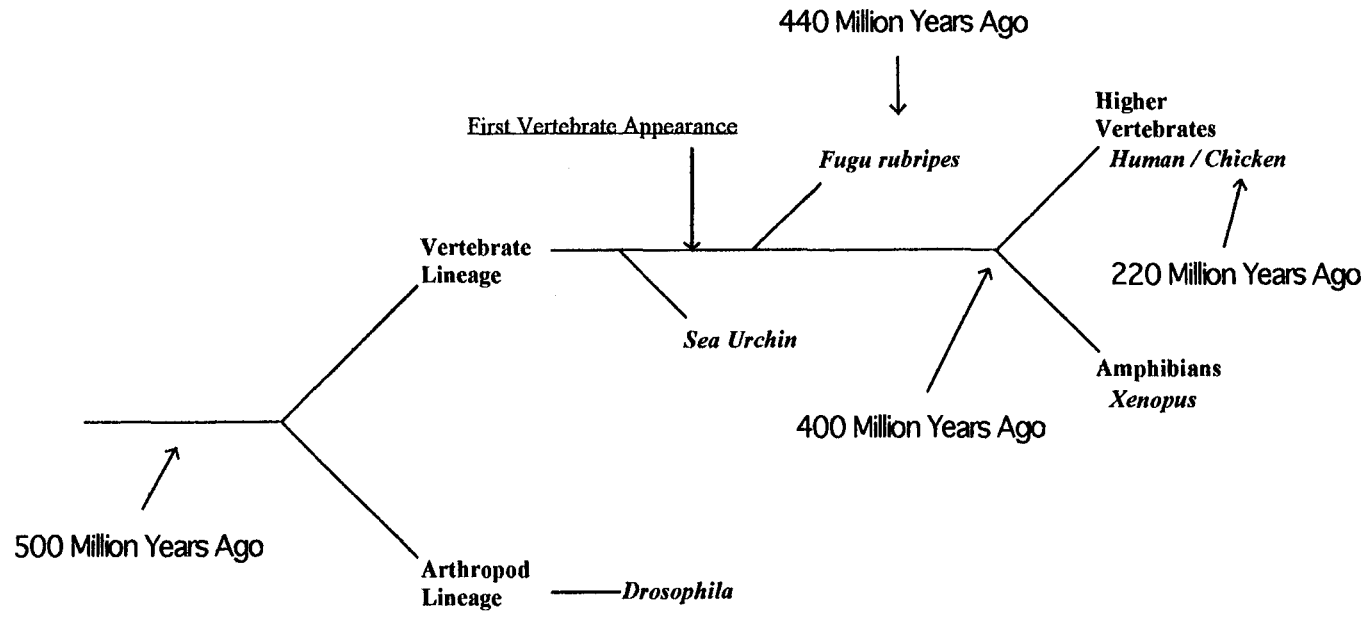
Table 2 **Current Predictions of Approximate Gene Number and Genome Size
in Organisms of Varying Evolutionary Lineage**

The approximate genome size and gene content is given for each organism. The mouse, human and pufferfish information is highlighted in yellow.

		Genes	Genome Size in Megabases
Prokaryota	<i>Mycoplasma genitalium</i>	473	0.58
	<i>Haemophilus influenzae</i>	1760	1.83
	<i>Bacillus subtilis</i>	3700	4.2
	<i>Escherichia coli</i>	4100	4.7
	<i>Myxococcus xanthus</i>	8000	9.45
Fungi	<i>Saccharomyces cerevisiae</i>	5800	13.5
Protoctista	<i>Cyanidioschyzon merolae</i>	5000	11.7
	<i>Oxytricha similis</i>	12000	600
Arthropoda	<i>Drosophila melanogaster</i>	12000	165
Nematoda	<i>Caenorhabditis elegans</i>	14000	100
Mollusca	<i>Loligo pealii</i>	>35000	2700
	<i>Ciona intestinalis</i>	Not Avail	165
Chordata	<i>Fugu rubripes</i>	70000	400
	<i>Danio rerio</i>	Not Avail	1900
	<i>Mus musculus</i>	70000	3300
	<i>Homo sapiens</i>	70000	3300
Plantae	<i>Nicotiana tabacum</i>	43000	4500
	<i>Arabidopsis thaliana</i>	16000 - 33000	70 - 145

Figure 2 Timeline of Vertebrate Evolution

Indicated on this timeline is the vertebrate/arthropod divergence approximately 500 million years ago. The emergence of the pufferfish and chicken are indicated at 440 million and 220 million years respectively.



observed through comparisons, identified conserved enhancer sequences. When these sequences were deleted from the mouse promoter, the deletion resulted in tissue-specific loss of activity when assayed in transgenic mice. In an additional study, the parallel sequences from the pufferfish promoter were found to reconstitute tissue-specific expression in a transgenic animal. The pufferfish genome has also been used successfully to identify regions of regulatory significance within the *Otx2* gene, which plays an essential role in skeletal patterning in mice (Kimura et al., 1997). In this instance, two cis-acting elements critical for directing *LacZ* transgene expression in the developing mouse embryo were isolated from the mouse promoter. Another example of the use of the pufferfish genome as a model for the identification of non-coding regulatory elements includes the discovery of a conserved 110 base-pair cis-acting element governing *Wnt-1* expression in the mouse neural plate (Rowitch et al., 1998). Sequences governing expression of the intercellular retinoic acid binding protein CRABP-1 have also been identified through use of sequence alignment involving the pufferfish genome (Kleinjan et al., 1998).

Instances of conservation of chromosomal synteny have also been found by comparison of the pufferfish and human genomes. In particular, the gene organization directly surrounding the pufferfish *c-Fos* gene has been found to mimic the gene sequence that is found on human chromosome 14 at the familial Alzheimer's disease (AD3) locus (Trower et al., 1996). In light of this body of research, it is clear that the pufferfish genome is an attractive system within which to search for sequences of regulatory significance since non-coding, unconstrained sequences will have had a maximum time with which to randomize by mutational events (Brenner et al., 1993). Through a comparison of promoters that span the evolutionary timeline (Figure 2) (pufferfish, chicken, mouse and human), insight into the conservation of regulatory elements within the PEA3 promoter may be gained. To accomplish this, a four-way promoter alignment will be constructed in order to facilitate the identification of regions within the promoter that are of importance to the regulation of PEA3 expression.

MATERIALS

The genomic chicken, and genomic Japanese Pufferfish (*Fugu rubripes*) libraries were obtained from Clontech Laboratories Inc., Palo Alto, CA, USA. Both the Hybond N nylon membranes and the (α - ^{32}P) dCTP used in the screening of these libraries were obtained from the Amersham Corporation, Oakville, ON, Canada. Ingredients used in the production of bacterial media (trypticase-peptone, yeast extract and granulated agar) were purchased from Becton Dickinson, Cockeysville, MD, USA. Dr. T. Jessell (Columbia University, New York, NY, USA) provided the 3' partial chicken PEA3 cDNA probe (cPEA3c.1) that was used in the initial screening of the chicken genomic cosmid library.

All restriction enzymes were purchased from Gibco BRL, Burlington, ON, Canada. Other modifying enzymes (large fragment of DNA polymerase I, *Taq* DNA polymerase, T4 DNA ligase), agarose, nucleotides (dATP, dGTP, dTTP, dCTP), random hexamers and DNA standards (1 kb ladder, 100 bp ladder, λ HindIII Ladder) were also obtained from Gibco BRL.

Unincorporated (α - ^{32}P) dCTP was removed from random hexamer probe preparations by passage through a Sephadex G-50 column (Pharmacia, Uppsala, Sweden). The two vectors used during cloning procedures, pBluescript KS II and pCR 2.1 TA cloning vector, were obtained from Stratagene (La Jolla, CA, USA) and Invitrogen (Carlsbad, CA, USA) respectively. RNA extraction from tissues (adult zebrafish purchased locally) was carried out using TRIzol® Reagent (GIBCO BRL). The resultant RNA was then used for first strand cDNA synthesis using Superscript™ reverse transcriptase, also obtained from GIBCO BRL.

All oligonucleotide primers used during DNA sequencing and polymerase chain reactions were synthesized by Dinsdale Gooden (The Central Facility of the Institute for Molecular Biology and Biotechnology, McMaster University, Hamilton, ON, Canada).

METHODS

Isolation and Characterization of Chicken PEA3 Cosmids

Library Screening

Firstly, the genomic chicken cosmid library was titered by serial dilution followed by plating on LB-agar-ampicillin plates (100 $\mu\text{g}/\text{ml}$). In this fashion the titer of the library was determined to be 10^8 cfu/ml. In order to ensure full representation of the chicken genome during the library screen, the cosmid library was redundantly plated such that ten genomes were present at the beginning of the screening process. By dividing the chicken genome size of 1.2×10^9 bp (Burt et al., 1995) by the average insert size of the cosmid library (38 kb), it was determined that approximately 3.2×10^4 colonies would be required to screen the library once. Therefore, 3.2×10^4 colonies were plated on each of ten 132 mm nylon filters, on 150 mm LB-agar-ampicillin plates. These ten filters would then serve as master filters during subsequent screening steps. These plates were then incubated inverted at 37°C overnight until the colonies were observed to be approximately 1mm in diameter.

To aid in the selection of positive colonies following probe hybridization, replica filters (two) were produced for each of the ten 132 mm nylon master filters. The master filters were removed from their 150 mm LB-agar-ampicillin plates and placed on 3mm Whatman paper. A fresh replica 132 mm nylon filter was first pre-wetted by placement onto a LB-amp plate. This wetted filter was then placed on top of the master nylon filter and covered with a piece of 3mm Whatman paper and a glass plate. Pressure was then applied to ensure even transfer of bacterial colonies onto the replica filter. The glass plate and Whatman paper were then removed and a series of holes were made in the master/replica filter stack to aid in realignment following hybridization. A second replica filter was then produced in the same fashion.

Each of the resultant replica filters were then incubated overnight at 37°C on 150 mm LB-agar-ampicillin plates.

The replica filters were then prepared for probe hybridization in the following fashion. Firstly, each filter was placed colony side up on 3mm Whatman paper that had been pre-wetted with 0.5 N NaOH for 30 seconds. Secondly, the filters were transferred to a piece of 3mm Whatman wetted with 1 M Tris-HCl (pH 7.6) and allowed to soak for 30 seconds. Thirdly, filters were placed on Whatman paper soaked in 1 M Tris-HCl (pH 7.6)/1.5 M NaCl and allowed to stand for 30 seconds. Each replica filter was then immersed in 1 M Tris-HCl (pH 7.6)/1.5 M NaCl and gently agitated to remove loose bacterial debris. Filters were then placed on a fresh sheet of 3mm Whatman paper and allowed to dry.

DNA was then crosslinked to the nylon filters by exposure to UV light. Each filter was placed into a Stratalinker™ UV crosslinker and exposed to 120 000 μ J of UV energy. Following crosslinking, the replica filters were then immersed in 1 M Tris-HCl (pH 7.6)/1.5 M NaCl a second time and gently washed with a gloved hand to remove remaining bacterial debris. The filters were then allowed to dry on 3mm Whatman paper.

Replica filters were then prepared for probe hybridization by immersion and incubation in pre-hybridization solution. Aqueous pre-hybridization solution was prepared to the following proportions: 5x Denhardt's solution (100x Denhardt's solution: 10 g Ficoll 400, 10 g polyvinylpyrrolidone, 10 g BSA [bovine serum albumin] in a total volume of 500 ml), 5x SSC, 1.5% SDS, 0.1 mg/ml sheared salmon sperm DNA in a total volume of 200 ml. The pre-hybridization solution was preheated at 75°C prior to adding of the sheared salmon sperm DNA, which was boiled for 10 minutes before its addition. Replica filters containing crosslinked DNA were then incubated in the resultant solution for 2 hours at 65°C. Following this period the pre-hybridization solution was changed as it often became filled with residual bacterial debris (incubation of probe with replica filters in hybridization solution containing bacterial debris results in high levels of background following autoradiography).

Radiolabelled probe synthesis was performed using random hexamer oligonucleotide priming (Feinberg & Vogelstein, 1983). The DNA of interest (a 3' chicken PEA3 partial cDNA [cPEA3c.1] or an exon 6 genomic fragment), that was to be used as a probe, was prepared by restriction digestion or polymerase chain reaction. This DNA was then purified by agarose gel electrophoresis. The DNA fragment was then extracted from the agarose gel by Qiagen Gel Extraction™ using a microcentrifuge. The DNA fragment (100 ng) was then combined with 5 μ g of random hexanucleotide in a total volume of 10 μ l. This DNA mixture was then boiled for 5 minutes and then placed on ice. Following a 5 minute incubation on ice, 20 μ l of the following mixture: 3 μ l each of 10mM dATP, dTTP, dGTP; 3 μ l 20 mM DTT; 3 μ l 10x RP buffer [90 mM HEPES (pH 6.6), 100 mM $MgCl_2$]; 5 μ l 10 μ Ci/ μ l α -³²P dCTP was added to the DNA mix. In addition, 1 μ l (6 units) of Klenow fragment (large fragment of DNA polymerase I) was added to the resultant labeling reaction. This mixture was then incubated at 37°C for 2 hours. Following incubation, two 2 μ l aliquots of the probe reaction product were spotted on DE81 cellulose filters. These filters were then used for phosphate precipitation testing to determine the incorporation of radiolabelled nucleotides into the probe. One of the duplicate DE81 filters was washed with 0.5 M Na_2HPO_4 followed by rinsing with distilled water. The filter was then rinsed with 95% ethanol and allowed to air dry. The washed and unwashed filters were then placed in Beckman Ready-Safe Liquid Scintillation Cocktail™ in scintillation vials. The specific activity of the probes was then determined using a Beckman 6800 scintillation counter. Radiolabelled nucleotide incorporation was then calculated as a percentage by division of the washed specific activity by the unwashed specific activity. An incorporation percentage above 30% would indicate adequate radiolabelled nucleotide (α -³²P dCTP) incorporation into the probe. The remaining volume of the probe labeling reaction was then spun through a sephadex G-50 column to remove any unincorporated α -³²P dCTP. The column eluted probe was then boiled for 5 minutes, quenched on ice for 1 minute, then added to the prehybridization solution and incubated at 65°C overnight. Using the liquid scintillation results from the washed DE81 filter an estimated 500 000 probe counts/ml of hybridization solution was obtained.

Following overnight incubation with radiolabelled probe, the filters were washed at 65°C with two successive wash solutions. The filters were first washed at low stringency with 2x SSC/0.5% SDS for three consecutive fifteen minute periods. A second washing step was then carried out at higher stringency using 0.2x SSC/0.5% SDS for 30 minutes. Following the washing procedure the filters were then wrapped in saran wrap and taped to a backing which had Stratagene orientation markers fixed to it (these aid in the alignment of the autoradiogram with the replica filters). Filters and backings were then placed in autoradiography cassettes with Kodak XAR film with intensifying screens and exposed for one to three days at -80°C. Following exposure the autoradiography films were developed and matched to the filters using the Stratagene orientation markers and the holes that had been made in the filters during replica plating. Duplicate films were used to determine which colonies on the master filters represented a positive signal. These colonies were then selected by overlaying the master filters on top of the autoradiography film. Using a yellow pipette tip, it was possible to select a small number of colonies during the first round of screening, or a single colony prior to cosmid DNA preparation. During the primary and secondary rounds of screening the selected colonies, or colony, were placed into a 15 ml Falcon tube containing 1 ml of LB-ampicillin (100 µg/ml) media. The Falcon tubes were then vortexed to resuspend the bacteria evenly in the media. Serial dilutions of the original inoculate were then plated onto 132 mm nylon filters on 150 mm LB-agar-ampicillin plates. These plates were then incubated inverted at 37°C overnight. Serial dilution plates, which contained well-spaced bacterial colonies, were then selected for use during the subsequent screening round.

PCR Generation of a Chicken PEA3 Specific Genomic Probe

Following the initial library screening using the 3' partial chicken PEA3 cDNA probe (cPEA3c.1), a second screen was conducted using a genomic chicken PEA3 fragment containing exon 6. This screen was done to identify other cosmid inserts, which would encompass the entire chicken PEA3 locus. To

produce this probe, PCR (polymerase chain reaction) was performed using gene specific primers designed to intron sequences lying on either side of exon 6. The PCR reaction contained 100 ng of template DNA (cPEA3-B), 1 mM dNTPs, PCR buffer (20 mM Tris-HCl pH 8.4, 50 mM KCl), 5 U *Taq* DNA polymerase, 0.5 - 2.5 mM MgCl₂, 100 pmol of each primer (primer numbers AB. 12782 and AB. 12783). Each reaction was performed in a final volume of 100 μ l. Amplification of the product was carried out using the following PCR cycle: denaturation at 94°C for 30 seconds, primer annealing at 60°C for 30 seconds and primer extension at 72°C for 30 seconds. This cycle was repeated 30 times using a Perkin-Elmer GeneAmp PCR System 9600. Following amplification, 10 μ l of the PCR reaction volume was run on a 1% analytical agarose gel to visualize the product DNA.

Large Scale DNA Preparation

Following the library screening procedure, putative positive bacterial colonies were selected for cosmid purification using the Qiagen Plasmid Maxiprep Kit. As detailed above, single bacterial colonies were selected based on their position corresponding to a duplicate positive signal on replica autoradiography films. Once selected, these colonies were used to inoculate 500 ml volumes of LB-ampicillin (100 μ g/ml) media. Cosmid cultures were incubated for 16 hours at 37°C with constant shaking in a Fisher Scientific Controlled Environment Incubator Shaker. Cultures were then centrifuged at 4200 RPM at 4°C in a Sorvall RC-3B refrigerated centrifuge for 15 minutes. The bacterial pellets were then drained of all traces of supernatant. Pellets were then resuspended in 10 ml of resuspension buffer P1 (50 mM Tris-HCl, pH 8.0; 10 mM EDTA; 100 μ g/ml RNase A). Bacterial cells were then lysed with 10 ml of lysis buffer P2 (200 mM NaOH; 1% SDS) which was allowed to incubate at room temperature for 5 minutes. The bacterial lysate was then neutralized by addition of 10 ml of chilled neutralization buffer P3 (3.0 M potassium acetate, pH 5.5). Following addition of buffer P3, samples were incubated on ice for 20 minutes. Samples were then centrifuged at 4200 RPM at 4°C for 30 minutes. During the centrifugation

step, a Qiagen-tip 500 was equilibrated with 10 ml of equilibration buffer QBT (750 mM NaCl; 50 mM MOPS, pH 7.0; 15% isopropanol; 0.15% Triton X-100). Following centrifugation, the supernatant was applied to the Qiagen-tip 500 column. However, to avoid bacterial debris from entering the column, the sample supernatant was first passed through Miracloth (Calbiochem, La Jolla, CA.) before its addition. The sample was then allowed to drip through the column resin by gravity flow. Once the entire sample had drained through the column resin, two successive washes with wash buffer QC (1.0 M NaCl; 50 mM MOPS, pH 7.0; 15% isopropanol) were carried out. The DNA was then eluted from the column using 15 ml of elution buffer QF (1.25 M NaCl; 50 mM Tris-HCl, pH 8.5; 15% isopropanol). DNA was then precipitated by addition of 10.5 ml (0.7 volumes) of isopropanol to the column eluate. The sample was then mixed by inversion and then centrifuged at 9500 RPM at 4°C in a Beckman RC-5B refrigerated superspeed centrifuge. The supernatant was then decanted off of the DNA pellet and 2 ml of 70% ethanol was added. A second centrifugation at 9500 RPM was then carried out for 5 minutes. The supernatant was once more decanted from the pellet, which was then allowed to air-dry for 5-10 minutes. DNA pellets were then resuspended in TE (10 mM Tris-HCl, pH 8.0; 1.0 mM EDTA, pH 8.0) and quantified by UV spectrophotometry using a Beckman DU 640 Spectrophotometer.

Restriction Digestion of Putative Positive Cosmids

Following purification of putative positive cosmids from the genomic library screen, restriction endonuclease digestion followed by southern blot (Southern, 1975) analysis was performed. Restriction digestion of cosmid DNA was carried out by digesting 1 μ g of each DNA sample in a 20 μ l reaction volume according to the manufacturer's specifications. Digestion reactions were incubated at 37°C for 4 to 6 hours to ensure complete digestion of the cosmid DNA. The DNA fragments produced by restriction endonuclease digestion were separated by agarose gel electrophoresis in 0.8% agarose gels at 80 V.

Southern Blot Analysis of Cosmid DNA

After electrophoresis the gel was denatured by soaking for 45 minutes in several volumes of 1.5 M NaCl, 0.5 N NaOH with gentle agitation. The denatured gel was then briefly rinsed in several volumes of deionized water to remove any residual denaturing solution. The gel was then neutralized by soaking for 30 minutes in 1 M Tris-HCl (pH 7.4), 1.5 M NaCl at room temperature with constant agitation. Lastly, the gel was soaked in 10x SSC for 30 minutes. The DNA was then transferred to a Hybond N nylon membrane by capillary action (Sambrook et al., 1989). Following the transfer step the DNA was crosslinked to the nylon membrane by exposure to 120 000 μ J of UV energy using a Stratagene Stratalinker™ UV crosslinker. The nylon membrane was then soaked in 6x SSC for 5 minutes at room temperature with gentle agitation to remove any residual agarose. The membrane was then prehybridized and hybridized in the same fashion as described in the library screening protocol.

DNA Sequencing using Chicken PEA3 Specific Primers

Since a 3' partial chicken PEA3 cDNA sequence existed prior to isolation of cosmids bearing genomic chicken PEA3, it was possible to design sequence specific primers for use in DNA sequencing. Exon specific primers, as well as primers designed to newly found intron sequence, were used to sequence the region of chicken PEA3 spanning from exon 6 to exon 13. In order to sequence 5' from exon 6, a series of intron specific sequencing primers were synthesized and used in DNA sequencing reactions. In this fashion genomic sequence was obtained up to the intron residing between exons 4 and 5. The remainder of the genomic sequence was obtained through the sequencing of subcloned restriction digestion fragments (using a combination of T3, T7 and gene specific primers) from the 5' end of the PEA3 gene. DNA sequencing was performed by the Central Facility at the Institute for Molecular Biology and Biotechnology (MOBIX) at McMaster University using an ABI 373 automated DNA sequencer.

Subcloning of 5' Genomic Chicken PEA3 Restriction Digest Fragments

In order to facilitate the sequencing of genomic chicken PEA3 and its promoter, a number of restriction digestion fragments were subcloned. Knowing that the promoter region of mouse and human PEA3 has a high GC content, consistent with the presence of a CpG island (Barrett, 1997), the chicken PEA3 cosmid was digested with Not I (recognition sequence GCGGCCGC) and Xba I. The resultant fragments were then cloned into the pBluescript KS II vector (either as Not I/Xba I directional or Xba I non-directional clones). Cosmid DNA (1 μ g) was digested with Not I and/or Xba I according to manufacturer's specifications. This digestion was incubated at 37°C overnight to ensure complete digestion of cosmid DNA. Vector DNA was also digested with Not I and Xba I, or Xba I alone, for 1 hour at 37°C. To prevent religation of the vector plasmid that had been digested with Xba I alone, the DNA was subjected to dephosphorylation using calf intestinal alkaline phosphatase (CIAP). Following the restriction digestion with Xba I, 2 μ l of 10x dephosphorylation buffer (500 mM Tris-HCl pH 8.5, 1 mM EDTA) was added to the digestion reaction along with 5 U of CIAP enzyme (1 μ l). The dephosphorylation reaction was incubated for an additional 1 hour at 37°C. Both vector and cosmid insert DNA were then purified by electrophoresis at 100 V for 4 hours in a 0.8% agarose gel. Following electrophoresis, the 3 kb linearized pBluescript KS II band and the cosmid digestion bands were excised from the gel using sterile scalpels. The DNA was then removed from the gel slices using the Qiagen Gel Extraction Kit. A second agarose gel was then run to allow for verification of the gel extraction procedure as well as to allow for approximation of DNA amounts during ligation reactions to follow. Ligation reactions were then prepared such that insert DNA was present in molar excess to vector DNA based on approximation from the analytical agarose gel that was run following Qiagen DNA extraction. The ligation reaction consisted of 1 μ l of vector DNA, 1-6 μ l of insert DNA (based on approximation), 2 μ l of 5x ligation buffer (0.66 M Tris-HCl, pH 7.5, 50 mM MgCl₂, 50 mM DTT, 10 mM ATP) and 1 μ l (1 U) of DNA ligase (Gibco BRL). The volume of the ligation reaction was then adjusted to 10 μ l with de-ionized distilled water. The ligation was then incubated at room temperature for 16 hours.

Prior to transformation of the ligation reactions into bacteria, competent cells were produced. DH5 α bacteria (genotype - F ϕ 80*lacZ* Δ M15 Δ (*lacZYA-argF*)U169 *deoR recA1 endA1 hsdR17*(r_K⁻,m_K⁺) *phoA supE44 λ thi-1 gyrA96 relA1*) were grown in 10 ml of LB media without supplements (ampicillin etc.) at 37°C overnight with gentle shaking. These starter bacterial cultures were then used to inoculate 200 ml volumes of LB media and allowed to grow at 37°C until they reached an OD₆₀₀ (optical density at a wavelength of 600 nm) of 0.5 (approximately 2 hours of incubation was required). The cultures were then centrifuged at 4000 RPM for 10 minutes at 4°C. The supernatant was then removed from the bacterial pellet. At this point 40 ml of cold solution 1 (10 mM NaOAc pH 5.6, 50 mM MnCl₂, 5 mM NaCl) was used to resuspend the bacteria, and allowed to incubate on ice for 20 minutes. The resuspended volume was then centrifuged at 4000 RPM for 10 minutes at 4°C. The supernatant was once again removed and the pellet was resuspended in 4 ml of cold solution 2 (10mM NaOAc, 5% glycerol, 70 mM CaCl₂, 5 mM MnCl₂). The resultant resuspensions were then flash frozen in an ethanol/dry ice bath and stored at -80°C in 100 μ l aliquots.

Ligation reactions were then used to transform competent DH5 α bacteria. Competent cell aliquots were removed from -80°C and thawed on ice. Once thawed, 100 μ l aliquots of competent cells were placed in chilled 15 ml Falcon tubes. The entire 10 μ l ligation reaction was then added to the 100 μ l cell volume. Bacterial cells were then incubated on ice for 30 minutes. The cells were then heat-shocked at 37°C for 45 seconds. The Falcon tubes were then placed on ice for 2 minutes. Following recovery, 1 ml of room temperature LB (no additives) was added, and the transformed cells were incubated with gentle agitation for 1 hour at 37°C. Transformed cells were then plated on LB-agar-ampicillin (100 μ g/ml) plates, which had previously been spread with 40 μ l each of 40 mg/ml X-Gal (5-bromo-4-chloro-3-indolyl- β -D-galactoside) and 100 mM IPTG (isopropyl- β -D-thiogalactopyranoside) to allow for blue/white selection of positive ligation products.

White colonies were selected for small scale DNA preparation by picking with a sterile yellow pipette tip. This tip was then used to inoculate 5 ml of LB-ampicillin (100 μ g/ml) media for overnight growth at 37°C with gentle shaking. Small scale DNA preparations were then carried out using alkaline

lysis (Birnboim et al., 1983). The resultant DNA was then subjected to restriction digestion with Not I and Xba I or with Xba I alone. The products of the restriction digests were then separated on a 1% agarose gel by electrophoresis at 100 V.

Isolation and Characterization of Pufferfish ERM Bacteriophage λ Clones

Library Screening

Before screening of the pufferfish λ genomic library was initiated, host bacteria (K802 genotype: *galK2 galT22 hsdR2(r_K⁻, m_K⁺) lacY1 mcrA- mcrB⁻ metB1 mrr⁺ supE44*) were first prepared by inoculation of 50 ml of LB, supplemented with 0.2% maltose and 10 mM MgSO₄, with a single bacterial colony. The bacterial inoculation was then allowed to incubate overnight with shaking at 30°C. Cells were then spun down at 2000 RPM for 10 minutes. The supernatant was then removed and the pellet resuspended in 15 ml of 10 mM MgSO₄ without vortexing. This bacterial suspension was then diluted to an OD₆₀₀ of 0.5/ml with 10 mM MgSO₄. The λ library lysate was then titered by serial dilution in lambda dilution buffer (10x buffer: 1.0 M NaCl, 0.1 M MgSO₄·7H₂O, 1.0 M Tris-HCl pH 7.5) followed by incubation of the diluted phage with 200 μ l of prepared K802 host bacteria at 37°C for 15 minutes. Following incubation 3 ml of 50°C top agarose (LB media containing 0.7% agarose) was added and the resultant mixture poured onto a LB-agar plate supplemented with 10 mM MgSO₄. Plates were incubated overnight at 37°C and the resultant number of plaques was used to determine the titer of the library lysate. In this fashion the titer of the pufferfish λ genomic library was determined to be 1×10^9 pfu/ml.

In order to ensure full representation of the pufferfish genome during the library screen, the λ bacteriophage library was plated redundantly, such that the genome would be represented ten times at the beginning of the screening process. The size of the pufferfish genome has been reported to be approximately 4.0×10^8 bp (Brenner et al., 1993). Therefore, by division of the genome size by the average

size of the λ inserts (15 kb) it was determined that 2.6×10^4 plaques would be required to screen the pufferfish genome once. To plate the library redundantly, 6 LB-agar-10 mM MgSO_4 dishes were plated at a density of 5.0×10^4 pfu/plate. This plating was accomplished by incubation of 5.0×10^4 pfu with 600 μl of freshly prepared bacterial host in 6 separate 15 ml Falcon tubes. To each of these tubes, 6.5 ml of top agarose was added and the resultant mixture spread onto each of 6 separate LB-agar-10 mM MgSO_4 plates. These plates were then incubated overnight at 37°C. Following incubation the plates were then allowed to cool at 4°C to prevent the top agarose from sticking to nylon membranes in subsequent screening stages.

Nylon membranes were then placed on the top agarose layer and allowed to stand for 2 minutes. A series of holes were made in the membranes to allow for alignment of replica filters at later stages of the library screen. Following the 2 minute transfer, the membranes were submerged in denaturing solution (1.5 M NaCl, 0.5 M NaOH) for a further 2 minutes. Next, the membranes were submerged in neutralizing solution (1.5 M NaCl, 0.5 M Tris-HCl pH 8.0) and incubated for 5 minutes at room temperature. Filters were then rinsed for 30 seconds by submergence in 0.2M Tris-HCl pH 7.5/2x SSC. Following the final rinse the filters were allowed to dry on 3mm Whatman paper. Replica filters were prepared in exactly the same fashion except that the transfer time was extended to 5 minutes. The filters were then treated in the same fashion as was described in the screening process of the chicken genomic cosmid library. However, during the library screening process for pufferfish ERM, two separate probes were used. Initially a full-length mouse PEA3 cDNA was used to isolate a single λ phage clone, which contained sequences from exon 4 to exon 13. During this screen both the probe hybridization and subsequent wash steps were done at 55°C. The second probe that was generated by PCR off of the original phage clone was used with a hybridization and wash temperature of 65°C.

λ Bacteriophage DNA Purification

Host bacteria (K802) were grown for 16 hours at 30°C in 5 ml of LB media supplemented with 0.2% maltose and 10 mM MgSO₄. An individual plaque, which had been determined to positively hybridize to the radiolabelled probe during the library screen, was selected using a sterile pasteur pipette. The plaque was then placed in a drop of overnight bacterial culture in a 6 ml polypropylene Falcon tube. An additional 2 ml of LB-0.2% maltose-10 mM MgSO₄ was then added and the samples incubated overnight with shaking at 37°C. Following incubation 100 μ l of chloroform was added and mixed thoroughly. Samples were transferred to a 15 ml Falcon tube and incubated at 37°C for 15 minutes. The supernatant was then removed and stored at 4°C. One milliliter of overnight culture was then incubated for 5 minutes at room temperature with 15 μ l of the primary supernatant, which had been collected earlier. An additional 50 ml of LB-0.2% maltose-10 mM MgSO₄ was then added. The resultant inoculation was then split into two 50 ml Falcon tubes for overnight incubation at 37°C. Following overnight incubation, 250 μ l of chloroform and 0.75 g of NaCl was added to each tube and shaken at 37°C until the NaCl had fully dissolved. Tubes were then centrifuged at 4200 RPM for 15 minutes at 4°C. The phage supernatants were then removed and added separately to 2.5 g of PEG (polyethylene Glycol - one for each replicate) in a 50 ml Falcon tube. Samples were incubated at 4°C for 1 hour. Each tube was then centrifuged at 4200 RPM for 30 minutes at 4°C. The supernatant was then removed and the bacteriophage pellet was resuspended in 0.6 ml of SM buffer without gelatin (0.1 M NaCl, 50 mM Tris-HCl pH 7.5, 10 mM MgSO₄-7H₂O). To remove any contaminating bacterial genomic DNA, 2.5 μ l of Dnase I (10 μ g/ml) was added and the samples incubated at 37°C for 1 hour. To protect against the subsequent degradation of λ phage DNA, 2 μ l of DEPC was added to each sample. At this point the individual samples were split into two eppendorf tubes, and 0.6 ml of bacteriophage lysis buffer (1M Tris-HCl pH 8.5, 1% SDS, 0.1 M EDTA) was added to each tube. Lysis of the phage was carried out by incubation of the samples at 70°C for 5 minutes. Following this incubation, 300 μ l (0.25 volumes) of 5 M potassium acetate was added. Samples were then incubated on ice for 30 minutes. Eppendorf tubes were then spun in a microcentrifuge at 4°C for 15

minutes. One milliliter of supernatant from the centrifugation was transferred to a fresh microcentrifuge tube and 600 μ l (0.6 volumes) of isopropanol was added. Precipitated DNA was pelleted by centrifugation at room temperature for 15 minutes. The supernatant was then removed by aspiration. The DNA pellet was then resuspended in 500 μ l of TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA). A single phenol/chloroform extraction was then performed by addition of 250 μ l each of buffer saturated phenol and chloroform. The aqueous upper phase was then removed and transferred to a fresh microcentrifuge tube. The DNA was then reprecipitated by addition of 0.1 volumes (50 μ l) of 3 M NaOAc and 2 volumes (1 ml) of absolute ethanol. DNA was pelleted by centrifugation at 13 000 RPM in a microcentrifuge at 4°C. DNA pellets were then resuspended in 50 μ l of TE buffer. The DNA was then quantified by UV spectrophotometry.

Restriction Digestion of Putative Positive λ Bacteriophage Inserts

Restriction digestion of the putative positive λ bacteriophage clones was carried out as was described for the chicken PEA3 cosmids.

Southern Blot Analysis of λ Bacteriophage DNA

Southern blot analysis of λ Bacteriophage DNA was carried out as was described for the chicken PEA3 cosmids. However, probe hybridization and subsequent washing was carried out at either 55°C or 65°C depending upon which probe was being used (mouse PEA3 cDNA or pufferfish ERM PCR generated exon 4 fragment respectively).

PCR Generation of a Pufferfish ERM Specific Genomic Probe

Initial screening of the pufferfish λ genomic library produced a single clone containing pufferfish ERM sequences from exon 2 to exon 13 plus additional genomic sequence 3' of the ERM gene. A second screen was performed using a genomic pufferfish ERM fragment containing exon 4 as a probe. To produce this probe, PCR was carried out using gene specific primers designed to intron sequences lying on either side of exon 4. The PCR reaction contained 100 ng of template DNA (pERM-A), 1mM dNTPs, PCR buffer (20 mM Tris-HCl pH 8.4, 50 mM KCl), 5 U *Taq* DNA polymerase, 0.5 - 2.5 mM MgCl₂, 100 pmol of each primer (primer numbers AB. 12422 and AB. 12423). Each PCR reaction was performed in a final volume of 100 μ l. The PCR cycle which was used to amplify this region of the pufferfish ERM gene was as follows: denaturation at 94°C for 30 seconds, primer annealing at 65°C for 30 seconds and primer extension at 72°C for 30 seconds. This cycle was repeated 30 times using a Perkin-Elmer GeneAmp PCR System 9600. Upon completion of the PCR cycling, 1/10 of the reaction volume was run on a 1% agarose gel to visualize the product DNA.

Subcloning of DNA Fragments Containing Pufferfish ERM

To facilitate the sequencing of pufferfish genomic ERM, it was desirable to subclone the restriction digestion fragments which were shown by southern analysis to share sequence with either the mouse PEA3 cDNA or pufferfish ERM exon 4 genomic probe. Southern blot analysis of pERM-A had shown that a 3.8 kb Xba I fragment had hybridized to the mouse PEA3 cDNA probe. Restriction digestion was carried out on pERM-A by incubation of 1 μ g of DNA with Xba I according to manufacturer's specifications. Vector DNA (pBluescript KS II) was also cut with Xba I for 1 hour at 37°C. To prevent the recircularization of the vector DNA, CIAP enzyme was added. Following the restriction digestion with

Xba I, 2 μ l of 10x dephosphorylation buffer (500 mM Tris-HCl pH 8.5, 1 mM EDTA) was added along with 1 μ l (5 U) of CIAP. The dephosphorylation reaction was allowed to proceed at 37°C for an additional 1 hour. Both the phage DNA restriction digestion products and the vector digestion/dephosphorylation product were then electrophoresed at 100 V for 4 hours in a 0.8% agarose gel. The 3.8 kb Xba I fragment and the digested vector band at approximately 3 kb were excised with sterile scalpels. The insert and vector DNA fragments were then extracted from the agarose gel slices using the Qiagen Gel Extraction Kit. The products of the gel extraction were then visualized on a second agarose gel to confirm that the extraction was successful and to allow for estimation of DNA amounts during the ligation to follow. The ligation reaction consisted of 1 μ l of vector DNA, 5 μ l 3.8 kb insert DNA (based on approximation), 2 μ l of 5x ligation buffer (0.66 M Tris-HCl pH 7.5, 50 mM MgCl₂, 50 mM DTT, 10 mM ATP), 1 μ l (1 U) of DNA ligase and 1 μ l of sterile distilled water. The ligation reaction was incubated at room temperature for 16 hours.

A similar approach was taken to the subcloning of a 7.0 kb Xho I band from pERM-2D which had hybridized to the pufferfish ERM exon 4 genomic probe during southern analysis. Phage DNA (1 μ g) was digested with Xho I according to manufacturer's specifications. Vector DNA (pBluescript KS II) was also digested with Xho I followed by a dephosphorylation treatment with CIAP as was described previously. Both vector and insert DNA were purified by electrophoresis and extracted from gel slices using the Qiagen Gel Extraction Kit. The ligation reaction consisted of 1 μ l vector DNA, 10 μ l of insert DNA, 4 μ l of 5x ligation buffer (0.66 M Tris-HCl pH 7.5, 50 mM MgCl₂, 50 mM DTT, 10 mM ATP), 1 μ l (1 U) of DNA ligase and 4 μ l sterile distilled water. Once again the ligation reaction was incubated at room temperature for 16 hours. Both the 3.8 kb Xba I and 7.0 kb Xho I ligation reactions were used to transform competent DH5 α as was previously described for the subcloning of chicken PEA3 restriction fragments.

DNA Sequencing of Pufferfish ERM

Sequencing of pufferfish ERM began by using T3 and T7 primers to begin sequencing the 3.8 kb Xba I fragment of pERM-A. Additional primers were then designed to newly found sequence within this DNA fragment. In this fashion pufferfish ERM genomic sequence from exon 10 to 13 was completed. Sequencing 5' of the 3.8 kb Xba I fragment was also carried out by the design of sequencing primers to newly determined sequence. Phage DNA was used as template for sequencing using these primers. Primer walking towards the 5' end of the gene completed the genomic sequence up to the 5' end of the pERM-A insert. The 7.0 kb Xho I subcloned fragment was also used to verify the genomic sequence as well as to generate approximately 800 bp of sequence upstream of the putative pufferfish ERM start codon.

Reverse Transcriptase Polymerase Chain Reaction Isolation of Zebrafish PEA3

Total zebrafish RNA was acquired by TRIZOL RNA extraction from adult zebrafish. Single zebrafish were homogenized in 3 ml of TRIZOL Reagent. This volume was then increased to 5 ml by addition of a further 2 ml of TRIZOL. The homogenized samples were then allowed to incubate for 5 minutes at room temperature. A single milliliter of chloroform was then added and the tubes shaken vigorously by hand for 15 seconds. Samples were then incubated at room temperature for 3 minutes. The aqueous phase was then separated by centrifugation at 4000-RPM for 15 minutes at 4°C in a Sorvall RC-3B refrigerated centrifuge. Following centrifugation the colorless upper aqueous phase was extracted and transferred into a fresh tube. The RNA was then precipitated by addition of 2.5 ml of isopropyl alcohol. The precipitation was allowed to proceed at room temperature for 15 minutes. RNA was then pelleted by centrifugation at 4200 RPM at 4°C for 30 minutes. Following centrifugation the supernatant was removed and the RNA pellet was washed once with 5 ml of 75% ethanol (DEPC - Diethyl-pyro-carbonate - treated).

The samples were mixed by vortexing and centrifuged at 4200 RPM at 4°C for 15 minutes. Supernatants were then removed and the pellets allowed to air-dry for 10 minutes before being dissolved in DEPC treated sterile water.

First strand cDNA synthesis was then carried out using Superscript™ reverse transcriptase (Gibco BRL) in a 20 μ l reaction containing 1 μ g of total zebrafish RNA and 250 ng of oligo (dT)₁₈ primer. This mixture was first heated to 70°C for 10 minutes followed by a quick chilling on ice. Following a brief centrifugation to collect the sample at the bottom of the eppendorf tube, 4 μ l of 5x first strand buffer (250 mM Tris-HCl pH 8.3, 375 mM KCl, 15 mM MgCl₂), 2 μ l 0.1 mM DTT, and 1 μ l 10mM dNTP mix (10 mM each dATP, dGTP, dCTP, dTTP) was added. The sample was then incubated at 42°C for 2 minutes, followed by addition of 1 μ l (200 units) of Superscript™ reverse transcriptase enzyme. The first strand synthesis reaction was then incubated for 50 minutes at 42°C. Following completion of the incubation the enzyme was heat inactivated by incubation at 70°C for 15 minutes. RNA, which was complementary to the first strand cDNA, was then removed by treatment of the sample with 2 units of E. coli RNase H, which was incubated at 37°C for 20 minutes.

The product of the first strand cDNA synthesis was then used as template for PCR amplification of a zebrafish PEA3 cDNA. In a 100 μ l reaction volume the following components were combined: 10 μ l of 10x PCR buffer (200 mM Tris-HCl pH 8.4, 500 mM KCl), 1-5 μ l 50 mM MgCl₂ (the optimum concentration of MgCl₂ must be determined for each template-primer pair), 2 μ l 10 mM dNTP mix, 2 μ l of each zebrafish PEA3 specific primer (10 μ M), 2 μ l cDNA (from first strand reaction), 1 μ l *Taq* DNA polymerase (5 units). The PCR reaction volume was then adjusted to 100 μ l by addition of sterile distilled water. The PCR cycle, which was then used to amplify the cDNA target, was as follows: denaturation at 94°C for 30 seconds, primer annealing at 65°C for 30 seconds and primer extension at 72°C for 60 seconds. This cycle was repeated 30 times using a Perkin-Elmer GeneAmp PCR System 9600. The PCR product was then analyzed by gel electrophoresis in a 1% agarose gel.

Based on the published sequence of zebrafish PEA3 (Brown et al., 1998), the expected size of the PCR product was 1.7 kb. A 1.7 kb product was visible on the ethidium bromide stained agarose gel. This

band was excised using a sterile scalpel and the DNA was extracted from the gel using the Qiagen Gel Extraction kit. This PCR product was then subcloned into the pCR 2.1 TA cloning vector. In a 10 μ l ligation reaction volume 3 μ l of gel extracted PCR product, 1 μ l 10x ligation buffer, 2 μ l pCR 2.1 vector, and 1 μ l T4 DNA Ligase (4 units) were combined. An additional 3 μ l of sterile water was then added to increase the reaction volume to 10 μ l. The ligation reaction was then incubated at 16°C for 16 hours. Transformation and selection of colonies for analysis was then performed in the same fashion as described earlier.

Isolation and Characterization of Pufferfish PEA3 Cosmids

Library Screening

The library screening procedure was identical to that described for the chicken genomic cosmid library screen. As was described during the screening of the pufferfish genomic λ phage library, the approximate size of the pufferfish genome is 4.0×10^8 bp (Brenner et al., 1993). The average insert size of the cosmid library used in this screen was 38 kb. It was therefore determined that 1.0×10^5 cfu would be required to represent the pufferfish genome once. To ensure that all cosmid inserts were present during the screen, the library was plated at ten fold redundancy, in that 1.0×10^5 cfu were plated onto 10 separate 150 mm LB-agar-ampicillin (100 μ g/ml) plates overlaid with 132 mm nylon membranes.

Two separate probes were used during the genomic pufferfish cosmid library screen for PEA3. The first probe was a full-length zebrafish PEA3 cDNA, which was isolated as previously described. The second probe that was used to isolate cosmids bearing genomic pufferfish PEA3 sequence was a 5' portion of the zebrafish PEA3 cDNA (exons 2 to 4). Each of these probes were hybridized, and the replica filters subsequently washed, at a temperature of 65°C.

A third library screen was conducted independently (from the effort that was described above) by Dr. M. Fried. A 636 bp genomic pufferfish PEA3 probe was produced by PCR (see PCR Generation of a Pufferfish PEA3 Genomic Probe to follow) from a cosmid (pPEA3-1) that was isolated by the full-length zebrafish PEA3 cDNA library screen. This probe was used to screen a gridded pufferfish genomic cosmid library that had been made available by the Human Genome Mapping Project Resource Centre (HGMP-RC).

PCR Generation of a 5' Zebrafish PEA3 cDNA Probe

Screening of the pufferfish genomic cosmid library using a full-length zebrafish PEA3 cDNA as a probe produced a single cosmid that contained genomic pufferfish PEA3 sequence from exon 5 to exon 13 inclusive. In order to obtain further 5' sequences, including the promoter, a second library screen would have to be performed using a different probe. The probe that was used was a PCR generated fragment of the zebrafish PEA3 cDNA spanning from exon 2 to 4. The PCR reaction contained 100 ng of template DNA (zebrafish PEA3 cDNA in the pCR 2.1 cloning vector), 1mM dNTPs, PCR buffer (20mM Tris-HCl pH 8.4, 50 mM KCl), 5 U *Taq* DNA polymerase, 0.5 - 2.5 mM MgCl₂, 100 pmol of each primer (primer numbers AB. 15392 and AB. 17059). Each reaction was performed in a final volume of 100 μ l. Amplification of the product was achieved by using the following PCR cycle: denaturation at 94°C for 30 seconds, primer annealing at 65°C for 30 seconds followed by an extension step at 72°C for 30 seconds. The above PCR cycle was repeated 30 times using a Perkin-Elmer GeneAmp PCR System 9600. Upon completion of the 30 cycles, 10 μ l of the total reaction volume was electrophoresed on a 1% agarose gel at 80 V.

PCR Generation of a Pufferfish PEA3 Specific Genomic Probe

A third probe was produced during the screening of the pufferfish genomic cosmid library. This probe was generated by PCR using primers (AB. 16819 and AB. 16372) designed to intron sequences lying on either side of exon 5 in the newly found pufferfish PEA3 genomic sequence. Using the reaction conditions described earlier (in PCR Generation of a 5' Zebrafish PEA3 Probe), and the initial pufferfish PEA3 cosmid (pPEA3-1) as a template, a 636 bp product was obtained. This DNA fragment was then sent to Dr. M. Fried at the Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London England, WC2A 3PX. The 636 bp genomic fragment was then used as a probe to screen a gridded pufferfish cosmid library.

5' RACE using a Pufferfish PEA3 Exon 5 Primer

Using the sequence of exon 5 from pufferfish PEA3 a gene specific primer (AB. 17060) was designed in an attempt to determine the nucleotide sequence of the 5' end of the cDNA. This primer was then sent to Dr. M. Fried where a 5' RACE (Rapid Amplification of cDNA Ends) experiment was carried out on poly(A)⁺ pufferfish heart RNA.

Large Scale DNA Preparation

Once the library screening was complete, large scale DNA preparation of putative positive cosmids was carried out. This procedure was done as was described for the preparation of chicken cosmid DNA. DNA preparation of twelve cosmids that were received from Dr. M. Fried was carried out in the same manner, except that the antibiotic used in the overnight culture of bacteria was kanamycin at a concentration of 60 μ g/ml. Two subcloned XmaI restriction fragments (900 bp and 3.3 kb) cloned into the

pBluescript KS II cloning vector and a 5' RACE product cloned into the pGEM-Teasy cloning vector were sent from the Fried laboratory as 500 ng spots on 3mm Whatman paper. Upon receiving these DNA samples, the spots were cut out using a sterile scalpel. The paper circles were then placed into sterile eppendorf tubes and 50 μ l of 10 mM Tris-HCl pH 8.0 was added. The DNA was allowed to resuspend for 5 minutes at room temperature. Sample tubes were then flash spun in a microcentrifuge at 13 000 RPM. From the supernatant, 2 μ l (approximately 10 ng, based on a 50% recovery rate) was then used to transform competent DH5 α bacteria as previously described. Single colonies were then selected to inoculate large scale DNA preparations.

Restriction Digestion of Putative Positive Cosmids

The restriction digestion of putative positive clones from the pufferfish genomic cosmid library screens were conducted in the same manner as was described for restriction digestion of chicken PEA3 cosmids.

Southern Blot Analysis of Cosmid DNA

Southern blot analysis of the products of the pufferfish genomic cosmid library screen was carried out as was described for the chicken PEA3 cosmids. However, a number of probes were used to perform southern analysis on putative positive cosmid clones. These probes include: full-length zebrafish PEA3 cDNA, 5' zebrafish PEA3 cDNA (exons 2 - 4), 5' RACE product of pufferfish PEA3 (5' from exon 5, provided by Dr. M. Fried). All probes were hybridized to their respective membranes and washed at 65°C.

A separate southern analysis was also carried out in the Fried laboratory on the twelve cosmids that were isolated from the gridded pufferfish cosmid library. In this case the 636 bp genomic pufferfish PEA3 fragment was used as a probe.

Subcloning of Pufferfish PEA3 Restriction Digest Fragments

Restriction endonuclease digestion products that had been demonstrated to hybridize to any of the probes used during southern analysis were recognized as potential candidates for subcloning. These subclones would then be used to facilitate the sequencing of genomic pufferfish PEA3. The 4.8 kb BamHI fragment from pPEA3-1, the 6.5 kb BamHI fragment from pPEA3-1A, the 2.8 kb EcoRI fragments from pPEA3-1A, pPEA3-1B and 019G13, the 900 bp XmaI fragment from 019G13 and the 3.3 kb PspAI fragment from pPEA3-1 were all subcloned into the pBluescript KS II cloning vector as previously described in a 10 μ l ligation reaction. Transformation of competent bacteria and subsequent small scale preparation of DNA occurred as previously described.

DNA Sequencing of Pufferfish PEA3

A combination of direct sequencing of cosmid insert DNA as well as sequencing of subcloned restriction digest fragments produced pufferfish PEA3 genomic sequence. In each case T3, T7, M13 forward, and M13 reverse primers located within the vector DNA were used to initiate the sequencing effort. Once genomic sequence had been obtained, specific primers were ordered to produce further sequence data.

Computer Assisted Analysis of Genomic Sequence

Sequence Editing

All genomic sequence that was produced for each of the three genes under study in this work was assembled by use of the Sequence Editing and Analysis subprogram of the DNA Star Lasergene software package.

Multiple Sequence Alignment

The promoter sequences of mouse, human, chicken and pufferfish PEA3 were aligned by use of the Clustal alignment tool within the Multiple Sequence Alignment subprogram of the DNA Star Lasergene software package. Further, identification of pufferfish ERM was possible by alignment of the assembled amino acid sequences corresponding to open reading frames for this putative pufferfish Ets protein with the amino acid sequences from the mouse PEA3 group members. From this alignment a table of percent similarity and divergence was obtained.

BLAST (Basic local alignment search tool) Analysis of Sequence Data

Once sequence information had been produced for a given target DNA template, the Sequence Editing and Analysis subprogram of DNA Star was used to copy the sequence into the BLAST homepage at <http://www.ncbi.nlm.nih.gov/BLAST> (Altschul et al., 1990). Once here the sequence of interest was then compared against either the nucleic acid database (nr) or the protein database (Swiss Protein) to identify regions that were homologous to known sequence.

Identification of Consensus Transcription Factor Binding Sites within Promoter Sequences

The promoters of human, mouse, chicken and pufferfish PEA3, as well as pufferfish ERM were surveyed for the presence of consensus transcription factor binding sites through use of the MatInspector 2.2 search program at <http://transfac.gbf.de/cgi-bin/matSearch/matsearch.pl> (Quandt, et al., 1995). The Sequence Editing and Analysis subprogram of the DNA Star Lasergene software package was used to insert the promoter sequence of interest into the search utility at the above web address. The parameters of the search were set such that the core binding site similarity was 100% and the adjacent binding site similarity was 85%.

RESULTS

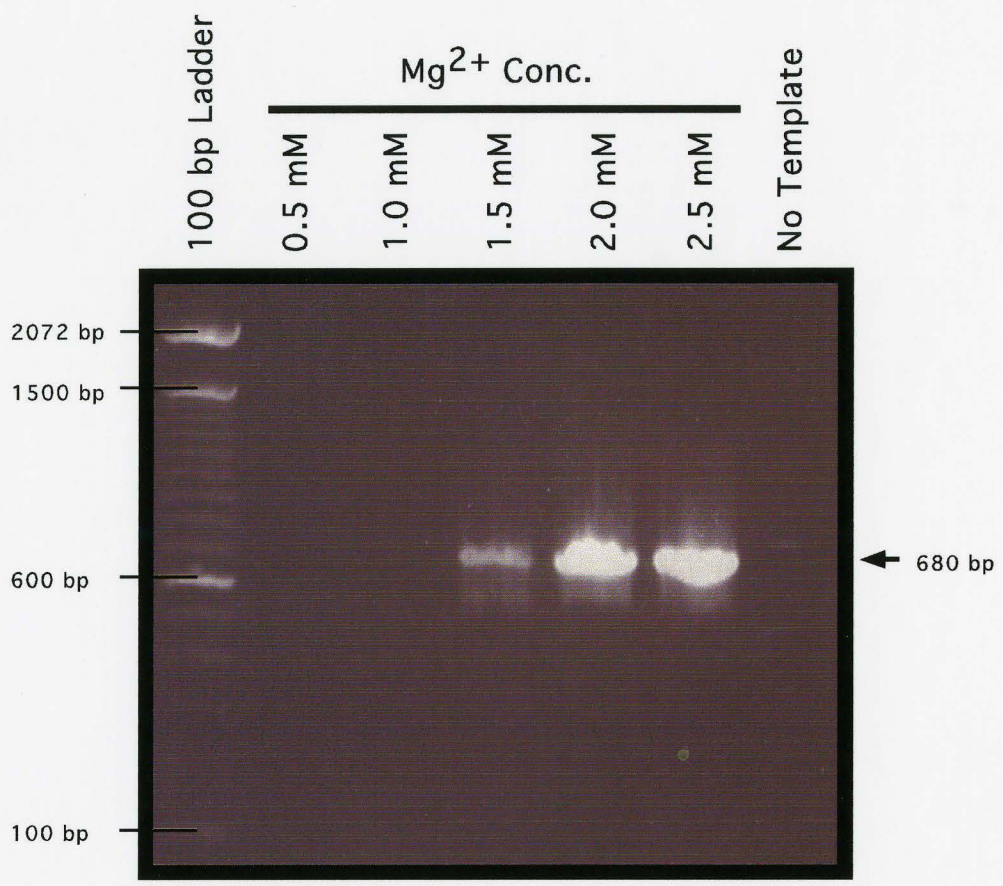
Isolation and Characterization of Chicken PEA3 Cosmids

Library Screening and PCR Generation of a Chicken PEA3 Specific Genomic Probe

In order to isolate and sequence the putative promoter region of chicken PEA3, a chicken genomic cosmid library was screened. Initially a 3' partial cDNA (cPEA3c.1 acquired from Dr. T. Jessell, Columbia University, New York, NY, USA) was used. This cDNA probe contained sequences inclusive of exons 7 to 13. Using this probe, an initial library screen produced six cosmids: cPEA3-A, cPEA3-B, cPEA3-C, cPEA3-D, cPEA3-E and cPEA3-F. Sequencing of cosmid cPEA3-B produced genomic sequence of the chicken PEA3 gene from exons 6 to 13. At this point a second library screen was performed in order to ensure that a number of different cosmid inserts were obtained, that when taken together would span the entire chicken PEA3 locus. To this end, gene specific primers were designed to intron sequences lying on either side of exon 6 (AB. 12782 and AB. 12783). Using these primers, PCR amplification of a 680 bp genomic fragment encompassing exon 6 was carried out using cPEA3-B DNA as a template for the reaction (Figure 3). PCR reactions were prepared at a number of differing $MgCl_2$ concentrations (0.5 mM - 2.5 mM), since the optimal Mg^{2+} concentration varies with each primer/template pair used. Using this specific genomic probe, three further cosmids were isolated: cPEA3-B1, cPEA3-B2 and cPEA3-B3.

Figure 3 PCR Generation of a Genomic Chicken PEA3 Specific Probe

PCR amplification of a 680 bp chicken PEA3 genomic fragment was carried out using gene specific primers that were designed to intron sequences lying on either side of exon 6 (AB. 12782 & AB. 12783). The template used in each of the PCR reactions was cPEA3-B. A Mg^{2+} titration was performed (0.5 mM - 2.5 mM) since the optimal concentration varies with each primer/template pair used. A no template negative control has also been included. A 100 bp DNA ladder has been included, and a number of size markers have been indicated. The position of the major PCR product (680 bp) has been noted.



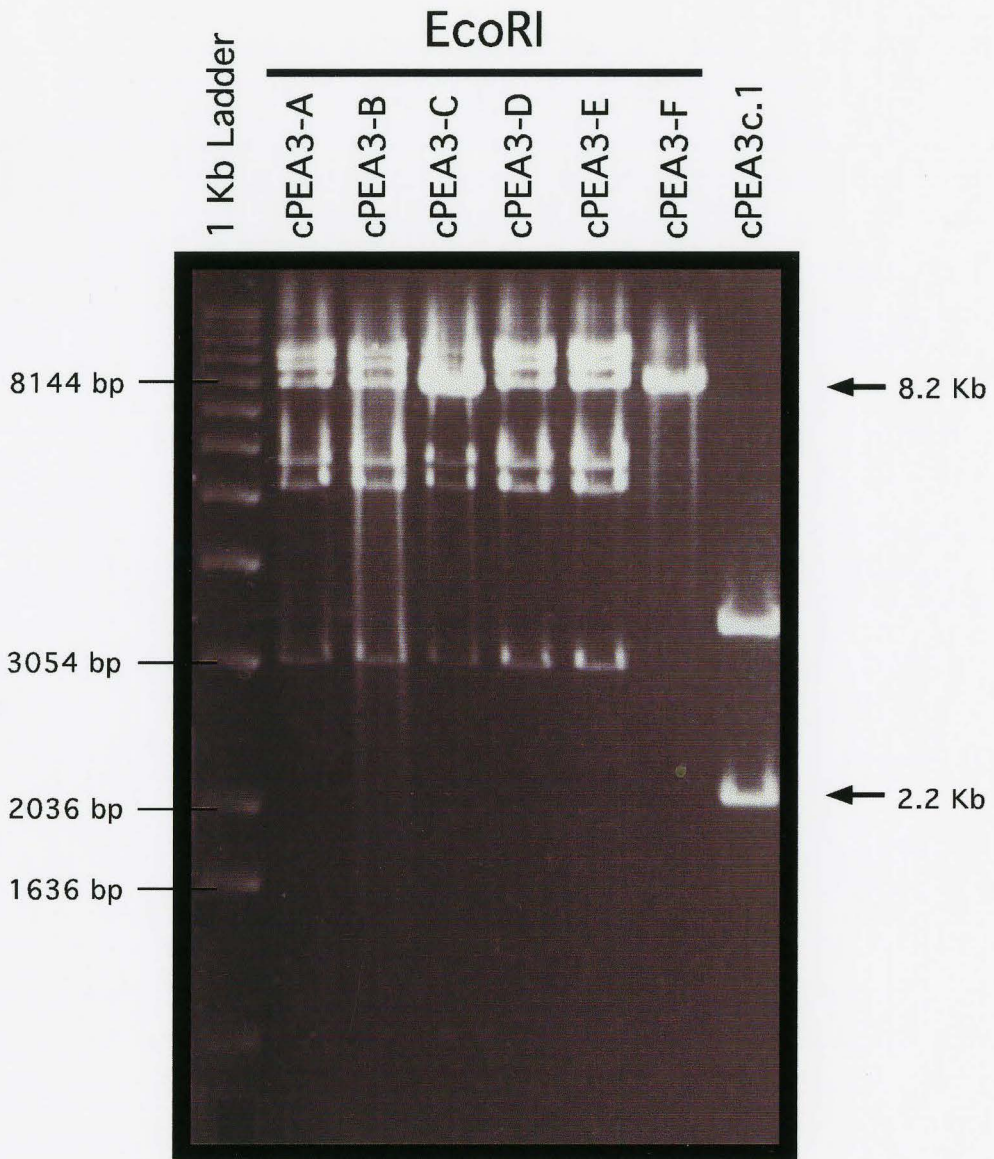
Restriction Digestion of Putative Positive Cosmids

Once the cosmid products from the initial screen of the chicken genomic library had been purified, they were subjected to restriction endonuclease digestion (EcoRI) followed by electrophoresis in a 0.8% agarose gel at 80 V. Analysis of the ethidium bromide stained gel revealed that cPEA3-A, cPEA3-B, cPEA3-C, cPEA3-D and cPEA3-E all possessed the same restriction digestion banding pattern when cut with EcoRI, while cPEA3-F possessed a unique digestion pattern (Figure 4).

Restriction endonuclease digestion was also performed on the cosmids that were produced from the second library screen using the genomic chicken PEA3 probe (Figure 5). When these three cosmids (cPEA3-B1, cPEA3-B2 and cPEA3-B3) were digested with EcoRI and the digestion fragments separated in a 0.8% agarose gel, it was clear that although each of the three new cosmids had unique restriction digestion patterns, cPEA3-B1 and cPEA3-B2 shared a number of EcoRI digestion fragments in common. A comparison of the two cosmids with cPEA3-B, one of the initially isolated cosmids also revealed shared digestion fragments. A band that migrated at an approximate size of 3.1 kb, relative to the 1 kb ladder, as well as a group of bands migrating at or above 8 kb appeared to be present in the EcoRI digestion products of cPEA3-B (used for comparison), cPEA3-B1 and cPEA3-B2. Given the fact that the banding pattern of cPEA3-B3 was found to be quite different from that found for the other cosmids it is likely that cPEA3-B3 was a false positive product, although Southern blot analysis of these cosmids would be required to validate this claim.

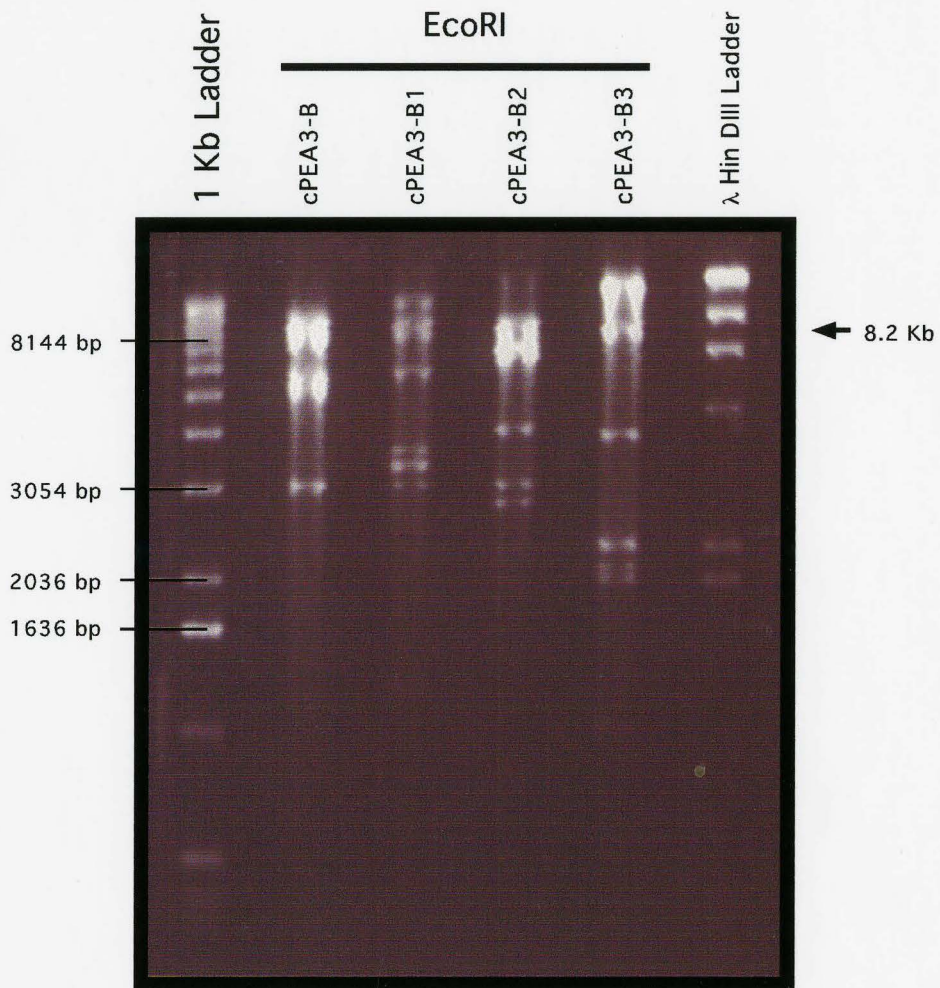
**Figure 4 Restriction Digestion of Putative Positive Chicken Cosmids
from the Initial Library Screen**

Putative positive cosmids cPEA3-A, cPEA3-B, cPEA3-C, cPEA3-D, cPEA3-E and cPEA3-F were digested with EcoRI and the product fragments separated by agarose gel electrophoresis. The partial chicken PEA3 cDNA cPEA3c.1 was also excised from the pBluescript II KS vector by EcoRI digestion and the products were electrophoresed to serve as a positive control during subsequent southern analysis. A 1 kb DNA ladder has been included, and a number of size markers have been indicated. The position of a common 8.2 kb restriction fragment within the six cosmid product lanes is shown. The position of the 2.2 kb restriction fragment containing the cPEA3c.1 partial cDNA is also noted.



**Figure 5 Restriction Digestion of Putative Positive Chicken Cosmids
from the Second Library Screen**

Putative positive cosmids cPEA3-B1, cPEA3-B2, and cPEA3-B3 were digested with EcoRI and the product fragments separated by gel electrophoresis. The chicken cosmid cPEA3-B was also digested with EcoRI and the product fragments electrophoresed to serve as a restriction fragment pattern comparison to the other three cosmids. This lane would also serve as a positive control during subsequent southern analysis using the cPEA3 genomic exon 6 probe that was produced using cPEA3-B as a PCR template. A 1 kb and a λ HindIII DNA ladder have been included, and a number of size markers have been indicated. The position of a common 8.2 kb restriction fragment has been noted.



Southern Blot Analysis of Cosmid DNA

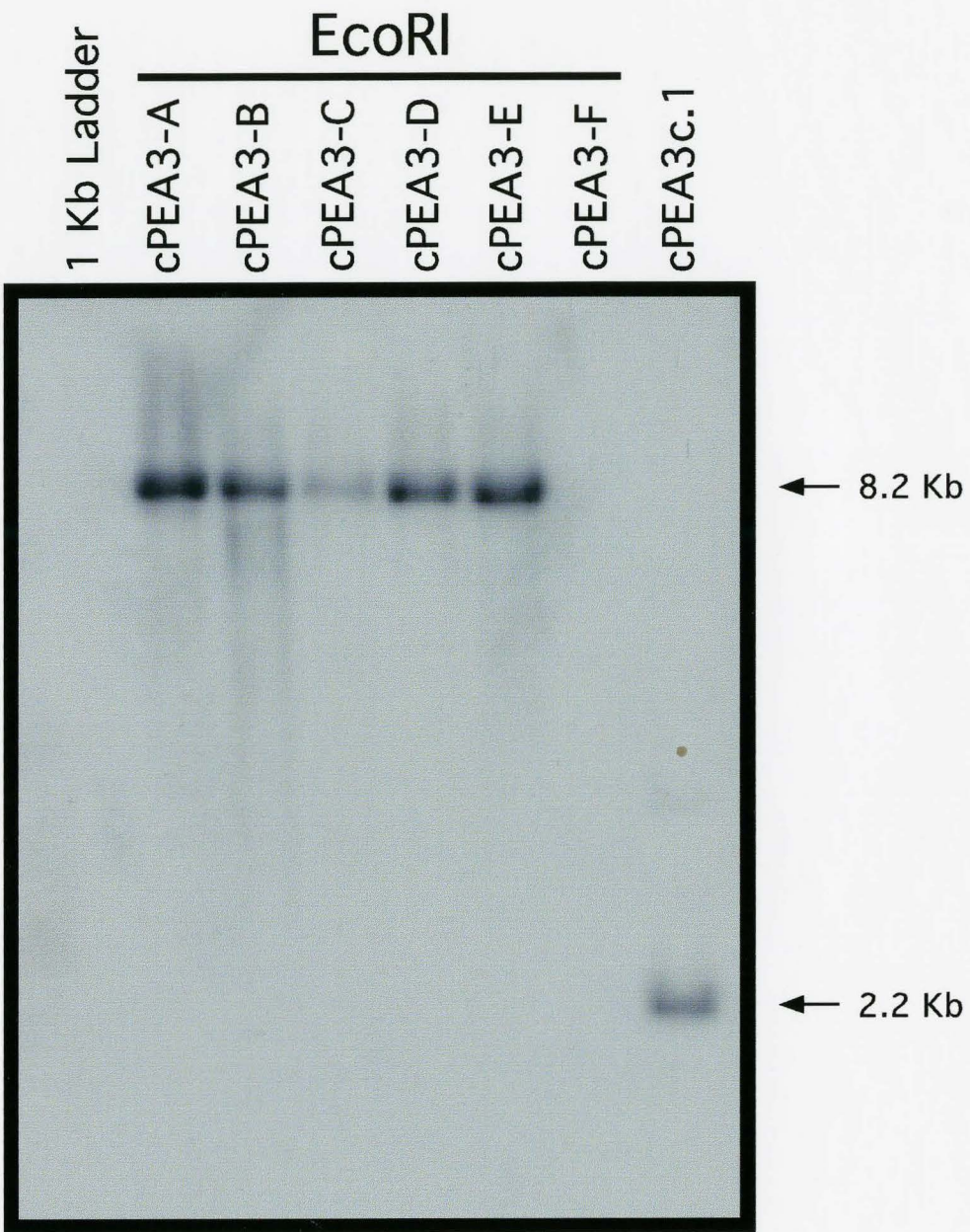
Subsequent to the restriction digestion of the cosmid clones from each library screen, southern blot analysis was performed. Following the electrophoresis of the restriction digestion products from each of the cosmids, the DNA bands were immobilized by transfer to nylon membranes. Hybridization of the 3' partial chicken PEA3 cDNA to the nylon membrane bearing the EcoRI digestion products of the six initial cosmid isolates was then carried out (Figure 6). This revealed that the probe hybridized specifically to a single restriction digestion band of an approximate size of 8.2 kb. This hybridization was observed in all six cosmids with the exception of cPEA3-F. This result, taken together with the restriction digestion patterns of the cosmids, suggests that the first five cosmids represented identical clones while cPEA3-F was a false positive.

Hybridization of the genomic chicken PEA3 probe to the immobilized EcoRI digestion products from cPEA3-B, cPEA3-B1, cPEA3-B2 and cPEA3-B3 showed probe hybridization to a band of an approximate size of 8.2 kb in the first three cosmids (Figure 7). As was the case with cPEA3-F, cPEA3-B3 did not hybridize to the probe and was therefore determined to be a false positive.

The combination of overlapping restriction digestion products and hybridization of the genomic probe to a common 8.2 kb digestion fragment confirmed that cosmids cPEA3-A to cPEA3-E share sequence in common with cosmids cPEA3-B1 and cPEA3-B2. It is important to note however, that the presence of unique digestion fragments in each of cPEA3-B1 and cPEA3-B2 demonstrates that these two cosmid inserts also possess unique genomic sequence.

**Figure 6 Southern Blot Analysis of Putative Positive Chicken
Cosmids from the Initial Library Screen**

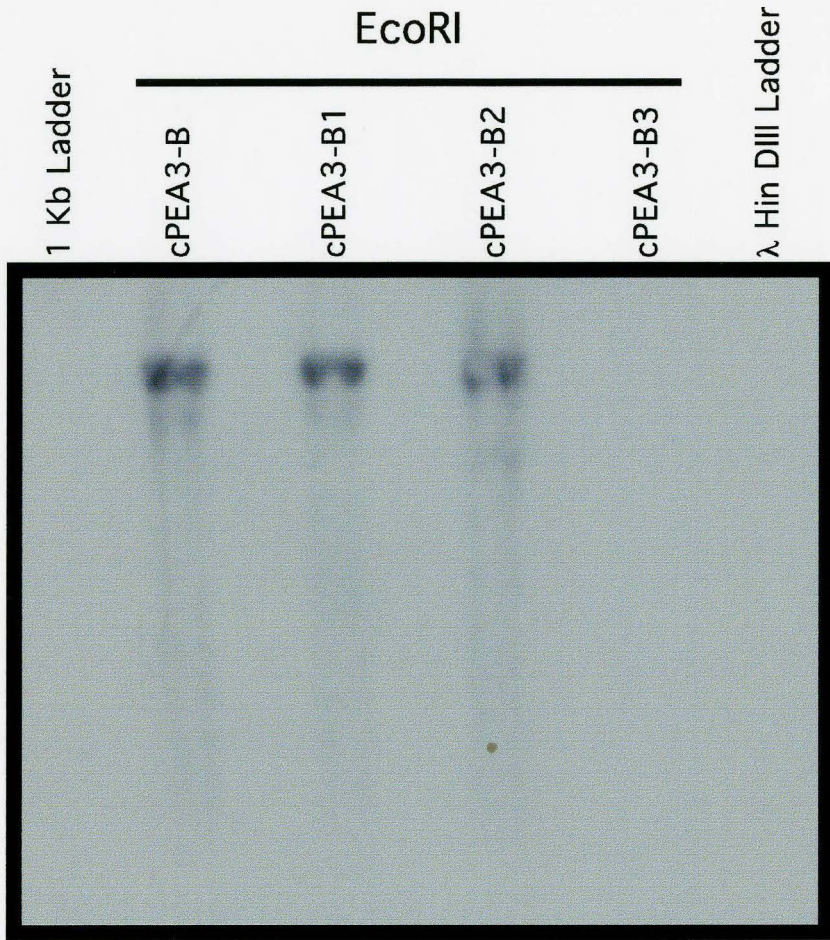
The 3' partial chicken PEA3 cDNA (cPEA3c.1) probe was hybridized to the immobilized EcoRI restriction fragments of cPEA3-A, cPEA3-B, cPEA3-C, cPEA3-D, cPEA3-E and cPEA3-F. The size of the restriction fragment (8.2 kb) that hybridized to the probe is indicated. The lane containing the immobilized 1 kb DNA ladder fragments is included. Further, the size of the restriction fragment in the cPEA3c.1 positive control lane (2.2 kb) that hybridized to the probe is shown.



**Figure 7 Southern Blot Analysis of Putative Positive Chicken
Cosmids from the Second Library Screen**

The 680 bp chicken PEA3 genomic probe containing exon 6 was hybridized to the immobilized EcoRI restriction fragments of cPEA3-B, cPEA3-B1, cPEA3-B2 and cPEA3-B3. The size of the restriction fragment (8.2 kb) that hybridized to the probe is indicated. The lanes containing the immobilized 1 kb and λ HinDIII DNA ladder fragments are noted.

8.2 Kb →



DNA Sequencing using Chicken PEA3 Specific Primers

Initially, cosmid cPEA3-B was used as a template for sequencing using exon specific primers, which were designed from the 3' partial chicken PEA3 cDNA sequence (cPEA3c.1). These exon specific primers reproduced the exon sequences of chicken PEA3 from exons 7 to 13 and also generated a large amount of intron sequence lying between these exons. Additional primers were then designed to intron sequences and used to produce continuous sequence spanning from slightly 5' of exon 6 to exon 13. All of the sequence information that was produced was joined together using the Sequence Editing and Analysis subprogram of the DNA Star Lasergene software package. The continuous 3' genomic sequence was then surveyed for the presence of useful restriction endonuclease recognition sites by using the Restriction Analysis and Mapping subprogram of the DNA Star package. From this analysis, restriction digestion followed by subcloning of the resultant fragments allowed for the sequencing of the 5' end of genomic chicken PEA3. Sequencing of the subcloned restriction digest fragments was carried out by using a combination of T3 and T7 sequencing primers followed by the use of gene specific primers designed from previous sequence data.

Subsequent sequencing of the cPEA3-B2 insert, using the T3 and T7 priming sites located within the pWE15 cosmid vector revealed that the 3' end of the insert resided approximately 500 bp 3' of the stop codon (exon 13) of chicken PEA3. Therefore, knowing that the minimum insert size of the cosmid library was 35 kilobasepairs, it was then reasonable to assume that the cPEA3-B2 cosmid insert encompassed the entire genomic chicken PEA3 gene. Based on this, cPEA3-B2 was used for all subsequent cloning experiments as well as the direct sequencing of the 5' end of the PEA3 gene. The complete sequence of genomic chicken PEA3 is presented in Appendix A.

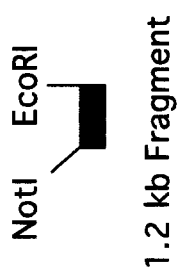
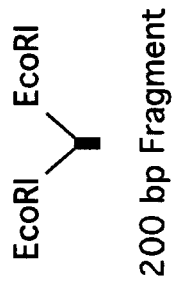
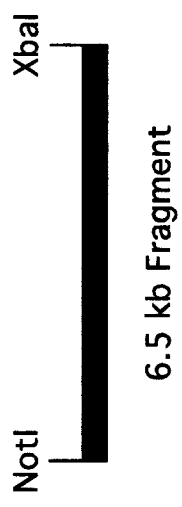
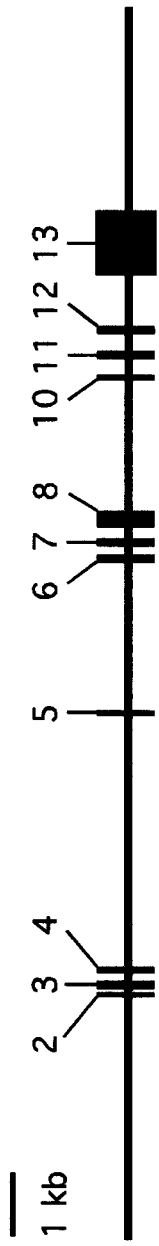
Subcloning of 5' Genomic Chicken PEA3 Restriction Digest Fragments

The genomic chicken PEA3 cosmid cPEA3-B2 was cut with the restriction endonucleases NotI and XbaI. From this digestion, eight fragments were observed following separation of the digestion products by electrophoresis (data not shown). A band was observed to migrate at an approximate size of 8.1 kb. Based on its size, it was concluded that this band, was the cosmid vector pWE15, which possessed NotI restriction sites at either end of the cosmid insert but did not possess XbaI sites internally. From the remaining restriction fragments observed on the ethidium bromide stained gel, five bands were excised, and the DNA extracted. The three largest bands (12, 9.0 and 7.5 kb) possessed NotI/XbaI produced sticky ends while the two smaller bands (3.8 and 3.0 kb) were the product of XbaI digestion. All five fragments were then subcloned into the appropriately digested pBluescript KS II vector. These subcloned products were then sequenced using the T3 and T7 priming sites located within the pBluescript vector sequence. Analysis of the sequence data revealed that the NotI/XbaI 12 kb and 7.5 kb DNA fragments possessed chicken PEA3 genomic sequence in the 5' end of the gene.

Further subcloning was performed from the NotI/XbaI 7.5 kb fragment by restriction digestion with NotI and EcoRI, followed by subsequent cloning into the pBluescript KS II vector. Two restriction digest products, a NotI/EcoRI 1.2 kb fragment and a 200 bp EcoRI fragment, were cloned. Sequencing of these restriction digest fragments confirmed a portion of the genomic sequence surrounding exon 4, which had been difficult to sequence in the 7.5 kb NotI/XbaI fragment. A summary of the subcloned chicken PEA3 restriction fragments is presented in Figure 8.

**Figure 8 A Summary of the Subcloned Restriction Digestion Fragments
from the Chicken PEA3 Cosmids**

The intron-exon organization of the chicken PEA3 gene is indicated to scale, with exons boxed in blue and introns depicted by thin black line. Each blue box has been numbered corresponding to which exon it represents. A 1 kb marker has been included to indicate scale. Below the depiction of the chicken PEA3 gene, the restriction fragments that were subcloned are indicated. On either end of the fragments the restriction enzyme recognition sites are indicated. The 12 kb XbaI fragment has not been drawn to scale (as indicated by the hatched line).



Reverse Transcriptase Polymerase Chain Reaction Isolation of Zebrafish PEA3

PCR amplification of zebrafish PEA3, using two sets of gene specific primers (AB. 15392/15395 & AB. 15393/15394 shown in Figure 9), was carried out on the product of first strand cDNA synthesis from total zebrafish RNA. Five separate PCR reactions were completed for each primer pair at various Mg^{2+} concentrations ranging from 0.5 mM to 2.5 mM. Following PCR, 10 μ l of the reaction volume was electrophoresed in a 1% agarose gel. The ethidium bromide stained gel revealed the presence of a major band at 1.7 kb in the 2.0 and 2.5 mM Mg^{2+} lanes in each PCR experiment (using different primer pairs) (Figure 10). Based on the published sequence (Brown et al., 1998), the expected size of the zebrafish PEA3 cDNA product was 1.7 kb. A number of other bands are also present at smaller sizes possibly due to non-specific hybridization of the primers to other cDNA targets. Incomplete extension of the 1.7 kb product could also explain the slight "laddering" effect that was observed on the gel. In each case a "no reverse transcriptase control", that did not contain first strand cDNA product, and a "no template control" did not exhibit product bands. Subcloning of the 1.7 kb band through the use of the TA cloning vector pCR 2.1, and subsequent sequencing using the M13 forward and M13 reverse sequencing primers, confirmed the sequence of the zebrafish PEA3 cDNA.

Figure 9 Zebrafish PEA3 cDNA Sequence

The 1710 bp zebrafish PEA3 cDNA sequence, as reported by Brown et al., 1998, is shown (Accession numbers: AJ003200, DRAJ3200). The position of nucleotides is indicated by numbering to the right of the sequence. Highlighted in green is the nucleotide sequence that encodes the acidic domain of zebrafish PEA3. Highlighted in yellow is the nucleotide sequence that encodes the Ets domain of zebrafish PEA3. Boxed in blue are in-frame termination codons that flank the open reading frame. The two in-frame start codons are underlined in black. The position of each of the zebrafish PEA3 cDNA sequence specific primers that were used to isolate the full length cDNA from total zebrafish RNA, and to generate a 5' cDNA probe, are indicated by black arrows with the primer numbers included below.

AB. 15392
AB. 15393
→

AGCCTAAATGACCACAAACGGT TAAGGCATGGATTATTAGATGGATGGATATCTGGACCA 60
 GCAAGTGCCTTATACTTTAGCTAATAGGTCGCAAGGAAATGGGCCCTAAATAGACTGTT 120
 GATGGCGACAAAAAGGAAATACATGGACGCAGAATTACCCCCTCAG GAATCTGAAGACCT 180
CTTTCAGGATTTAAGCCAACCTTCAGGAGACCTGGCTACCGAAGCTCAAGTTCCTGACAG 240

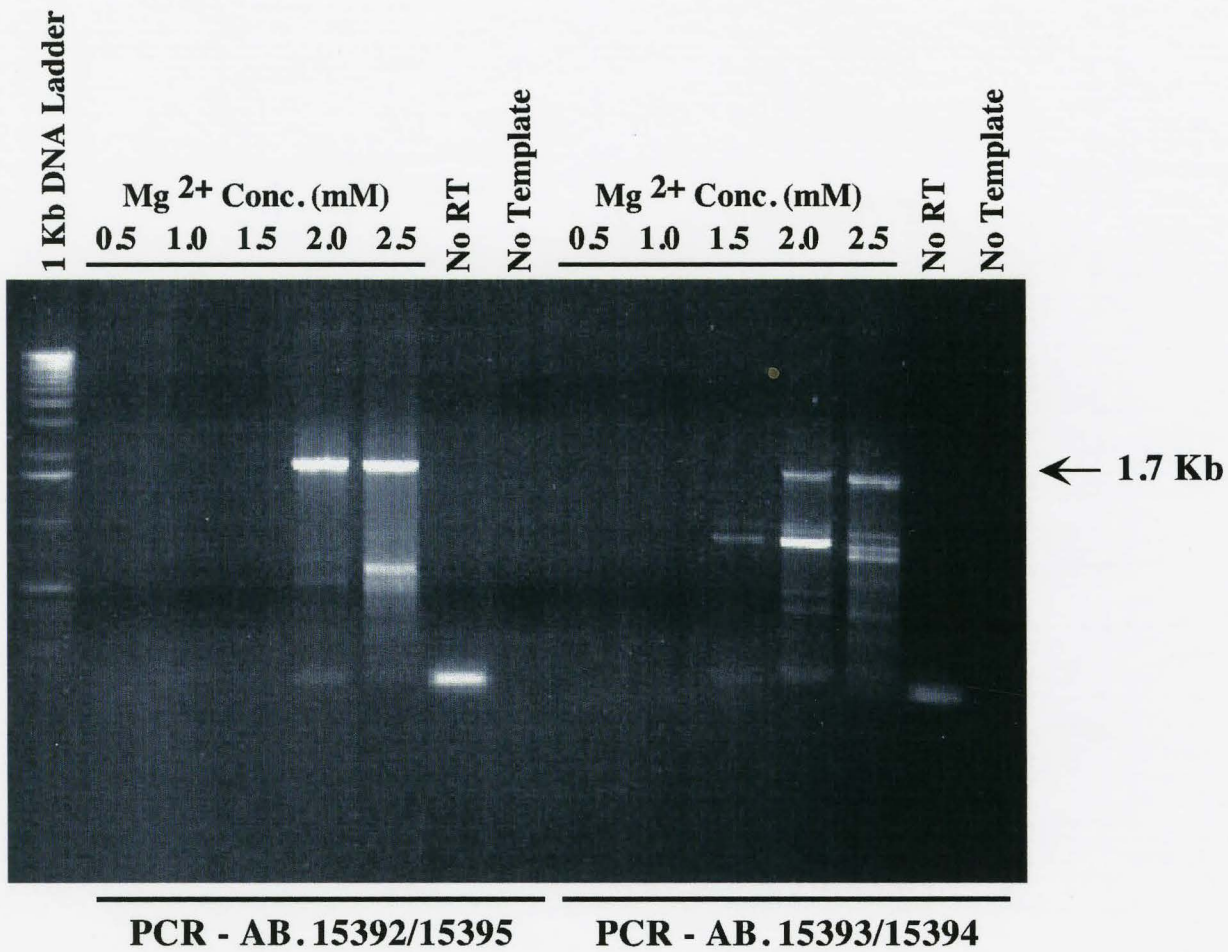
←
AB. 17059

CGATGAGCAGTTTGTTCCTGAC TTTCACTCAGAGAACTCAGTGGCATTCCACAGCCC GCC 300
 TGTGAAGATTAAGAAAGAGCCGCAGAGTCCCGGATCAGATCCCAGCCAGTCTGCAGCCA 360
 CAAGCAAAGCTTCAGCTACCCCAATGGAGAGCAGTGCCTTACGCCAGTGCCTATGAGCA 420
 GAAGAGAGCAGCCGTGGCAGGAGCAGGAGGATCTAAGAGCTCATGTCCAGCGACGCCAT 480
 GTCCCCTATGCAGCATTATCCCCCAAACCAACAGTGGGAACTCGGCAGGAATCAGGGTA 540
 TATGAATCCACCCTCATCCAGCCAATCCCACGCCTGCCACAGCCACAGTTACCCCATGAA 600
 CCCAGTTCAGGTTCCCTTCCGGCTCGGCAGAGATGTGCCACCATTGCTTCCCAAGG 660
 CCAGGCCCTGCAGCGCATCGATCCTGCCATGCCAGCGGAGGCGGAGGAGGCGGCTACCA 720
 TCGGCAACACTCTGACCCCTGCTTGCTTATATCCTCCTCAGCAGACCTTAAAGCAGGAGTA 780
 CATGGACCCTCTGTATGACCGAGCGGCCACATTAACGGACCACAGCCTCAGAGGTTTCC 840
 TCCTGCCATATGATGGTCAAACAGGAGCCCACCGACTACACTTACGAACCTGATGTGCC 900
 TGGCTGCCATCCATGTACCATCACAATGAAGGCTACTCCAACCCACAGCACAACAGTGA 960
 AGGCTACATGTTGAAAATGATTCCCGTGTGTGCCAGAGAAATTTGAAGGTGAGGTGAA 1020
 ACAAGAAGGGGGCAGTGTGTTTCGTGAAGGTGCCCCCTATCAGCGTCGT GGCTCACTTCA 1080
ACTATGGCAGTTCCTGGTTGCCCTTCTTGATGACCCAGCAATGCCACTTCATCGCATG 1140
GACGGGCCGTGGCATGGAATTCAAACTCATTGAGCCAGAGGAGGTGGCAAGACTCTGGGG 1200
GATGCAGAAGAACCGCCAGCCATGAACTATGACAAACTGAGTCGCTCTTTCGTTACTA 1260
TTATGAGAAGGGAATTATGCAAAAGGTGGCTGGTGAGCGTTATGTTTACAAATTTGTGTG 1320
TGAGCCAGAAGCTTTGATAACCTTGCTTTTCCCGACAATCAGCGGCCAAGTCTGAAGGC 1380
 GGAATTTGAGCGCTACGTCAACGAGGAGGACACTGTGCCACTGTCCCACCTCGATGAGGG 1440
 AGTTTCTTACCCTCCGGAACCAGCTGCCACCAACATGGGCCCTCAGCCCTACTCCAAAGG 1500
 CTACATGTACT TAATGGCACCGACATTCTCCTTTGTTTCCCATCCCAGTTTACCACCCGCT 1560
 GCACCTATTGCACATGCCACCCCATCCTCTTGATATAAGGCTATGTTTTATGTTTTG 1620
 TTTTGATTAATAATCTCATTAGGACTCTGCATTCTCCTCACTTCTGAGGCCTATCTA 1680
 ATCTAAACACAATACTGTGGTTCAAGAGATTGCTTA ← AB. 15394 1710

←
AB. 15395

Figure 10 Zebrafish PEA3 cDNA RT-PCR

PCR amplification of a 1.7 kb zebrafish PEA3 cDNA following first strand synthesis from total zebrafish RNA is shown. The left panel of the gel shows the results of PCR using zebrafish PEA3 specific primers AB. 15392 and AB. 15395. The right panel of the gel shows the results of PCR using zebrafish PEA3 specific primers AB. 15393 and AB. 15394. With each primer pair that was used a Mg²⁺ titration (0.5 mM - 2.5 mM) was carried out. PCR amplification was also performed on a no reverse transcription (RT) negative control. Further, a no template PCR negative control has also been included. A 1 kb DNA ladder has been included as a size marker. The position of the 1.7 kb major product is indicated.



Isolation and Characterization of Pufferfish PEA3 Cosmids

Library Screening and PCR Generation of 5' Zebrafish PEA3 cDNA and Genomic Pufferfish PEA3 Probes

In order to conduct a four-way alignment of the PEA3 promoter region from a variety of evolutionarily divergent organisms, a pufferfish genomic cosmid library was screened. This library was screened on two separate occasions using two different probes. The first probe was a full-length zebrafish PEA3 cDNA, which had been isolated by RT-PCR from total zebrafish RNA as described previously. Screening of the pufferfish genomic cosmid library with this probe produced a single cosmid isolate: pPEA3-1. Subsequent to the isolation of this cosmid, a second library screen was conducted using a PCR amplified fragment of the zebrafish PEA3 cDNA containing exons 2 to 4. A single primer (AB. 17059) was designed in exon 4 of zebrafish PEA3 and this primer was then used with one of the original primers (AB. 15392) to amplify the 5' end of zebrafish PEA3. Using the full-length cDNA as a template for the PCR reaction a single amplification product was produced at a size of 215 bp (Figure 11). Using this 5' zebrafish PEA3 probe, three further cosmids were isolated: pPEA3-1A, pPEA3-1B and pPEA3-1C.

Using a gridded cosmid library made available by the Human Genome Mapping Project Resource Centre (HGMP-RC), a third pufferfish genomic cosmid library screen was conducted using a genomic probe that was PCR amplified from pPEA3-1 (Dr. M. Fried and colleagues). Sequencing of pPEA3-1 generated genomic sequence spanning exon 5, including intron sequences adjacent to this exon. PCR primers, designed to intron sequences on either side of exon 5 (AB. 16819 & AB. 16372), were used to PCR amplify a genomic fragment of pufferfish PEA3. By using pPEA3-1 as a template for the PCR reaction, a single 636 bp product was produced (Figure 12), and this genomic fragment was subsequently used for the screening of a gridded pufferfish genomic cosmid library. This library screen produced 12 positive cosmid clones designated 019G13, 019G21, 037A22, 042K06, 073B12, 113F13, 117D03, 143G07, 184N11, 184M16, 196D02 and 196E02.

Figure 11 PCR Generation of a 5' Zebrafish PEA3 cDNA Probe

PCR amplification of the 5' end of the zebrafish PEA3 cDNA sequence was carried out using the primers AB. 17059 and AB. 15392. A Mg^{2+} titration was performed (0.5 mM - 2.5 mM) to determine the optimal concentration for this primer/template pair. A 100 bp ladder has been included and a number of the size markers have been indicated. The position of the 215 bp major band is noted.

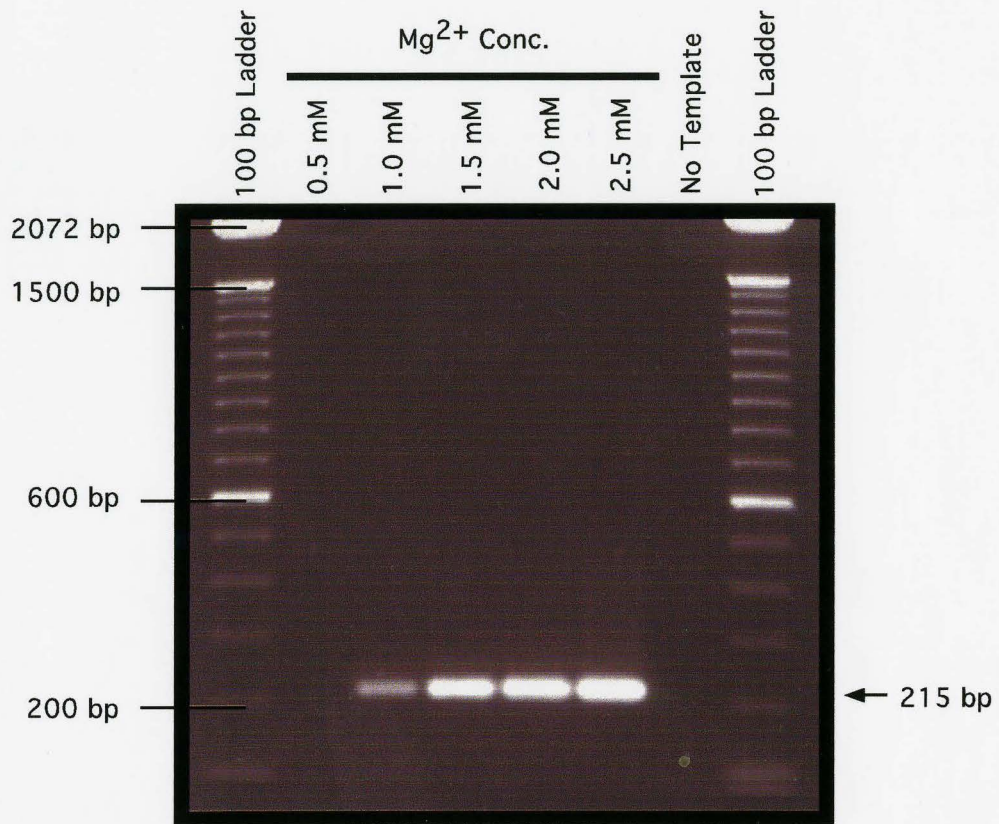
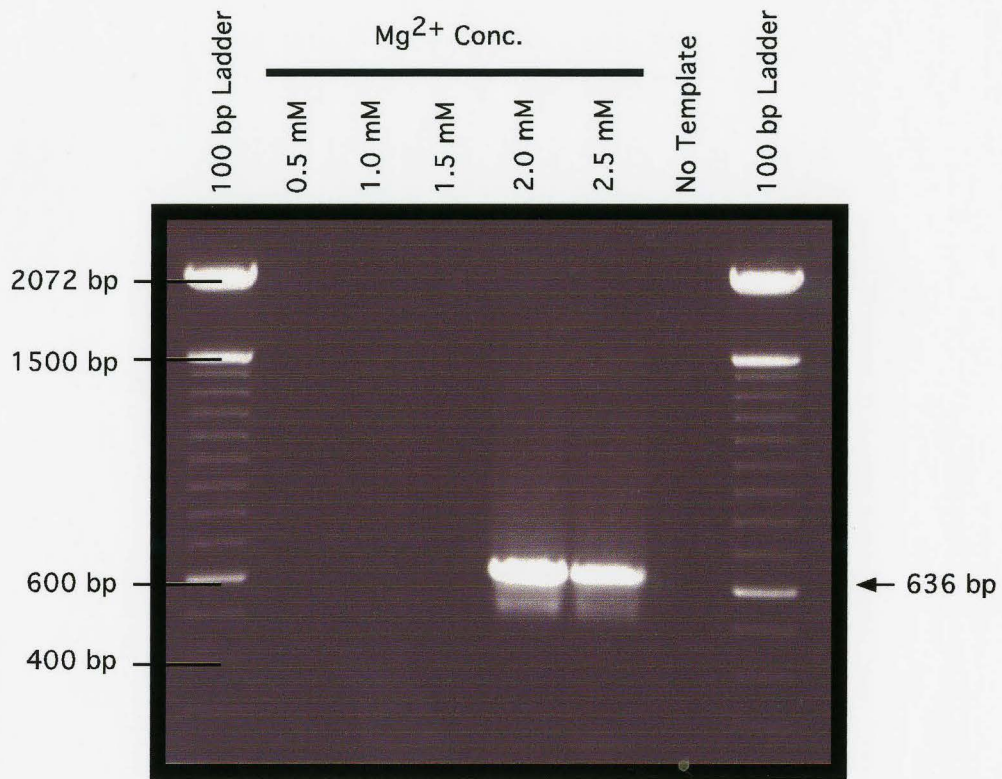


Figure 12 PCR Generation of a Genomic Pufferfish PEA3 Probe

PCR amplification of a 636 bp pufferfish PEA3 genomic fragment was carried out using gene specific primers that were designed to intron sequences lying on either side of exon 5 (AB. 16819 & AB. 16372). The template that was used in each of the PCR reactions was the pufferfish PEA3 cosmid pPEA3-1. A Mg^{2+} titration was performed (0.5 mM - 2.5 mM) since the optimal concentration varies with each primer/template pair used. A no template negative control has also been included. A 100 bp DNA ladder has been shown and a number of size markers have been indicated. The position of the 636 bp major PCR product has been noted.



5' RACE using a Pufferfish PEA3 Exon 5 Primer

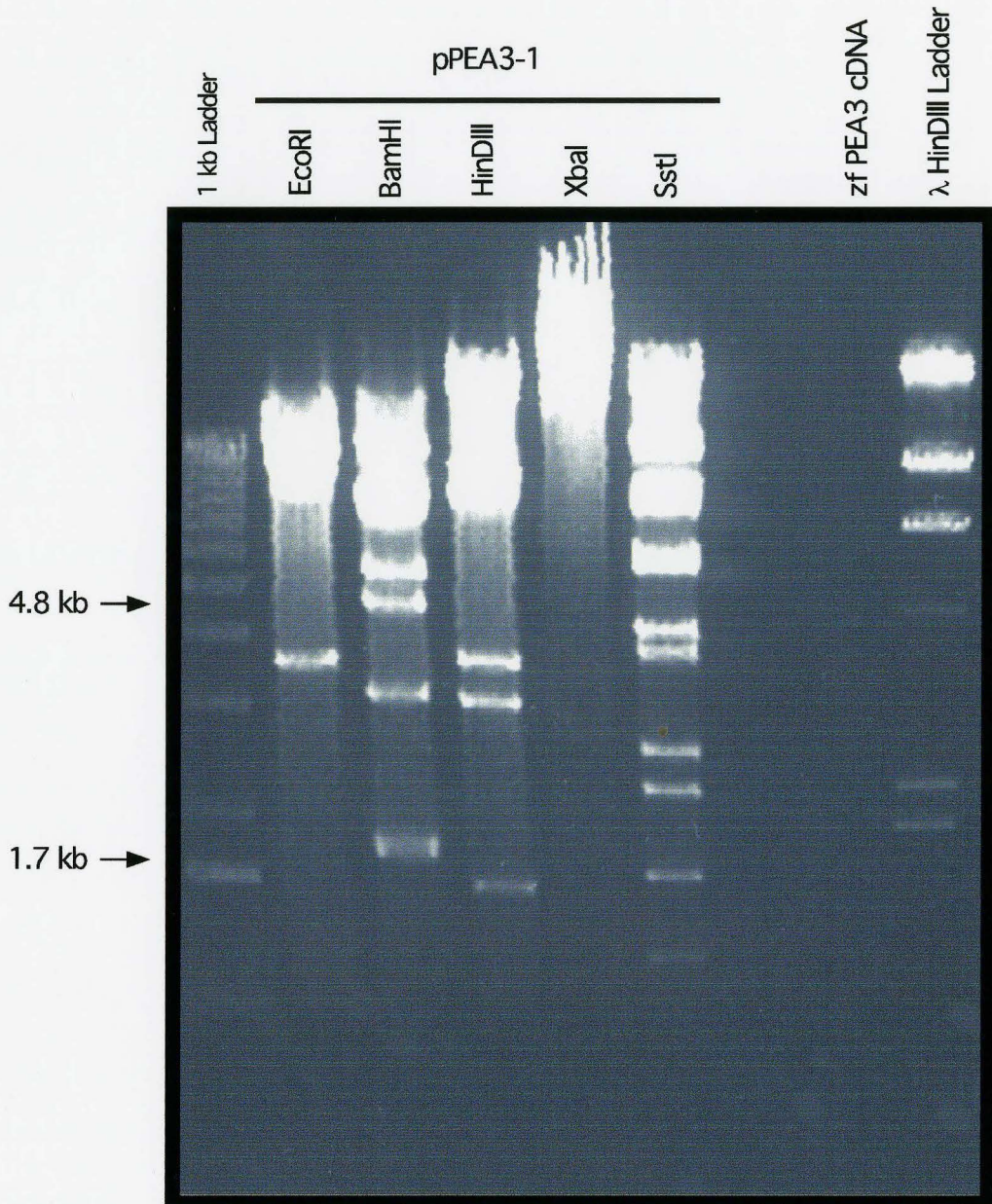
Using the sequence of the 636 bp genomic fragment from pufferfish PEA3, gene specific 5' RACE primers were produced. Using poly(A)⁺ RNA from pufferfish heart tissue, three products were produced using 5' RACE. The first product was 259 bp in length, and when translated, had an amino acid sequence that was extremely similar to zebrafish PEA3. The second 5' RACE product was 363 bp in length and possessed a nucleotide sequence that was identical to the first product, up to the 5' border of exon 3. Upstream from this point the two sequences were completely dissimilar, possibly because the 363 bp product retained intron sequence from a partially unprocessed mRNA. The third 5' RACE product was 334 bp in length and contained identical sequence to the first product with an additional 81 bp on the 5' end (See Appendix B for sequence). Presumably, this further sequence represents the 5' untranslated region (exon 1 or 1') of PEA3, possibly up to the transcriptional start site. The deduced amino acid sequence of the 334 bp product was aligned against the amino acid sequence of other PEA3 orthologues as well as other PEA3 group protein sequences and the highest degree of similarity was found to be shared with zebrafish PEA3.

Restriction Digestion of Putative Positive Cosmids

Following purification of the cosmid products from each stage of the library screen (including the 12 cosmids produced in the third library screen) restriction endonuclease digestion was performed. The initial library screening product pPEA3-1 was digested with EcoRI, BamHI, HinDIII, XbaI and SstI and the products were then separated by electrophoresis in a 0.8% agarose gel (Figure 13). Cosmids pPEA3-1A, pPEA3-1B and pPEA3-1C, produced from the second library screen, were digested with NotI, EcoRI, BamHI, XhoI, XbaI and HinDIII. The products of these digestions are shown in Figures 14 and 15.

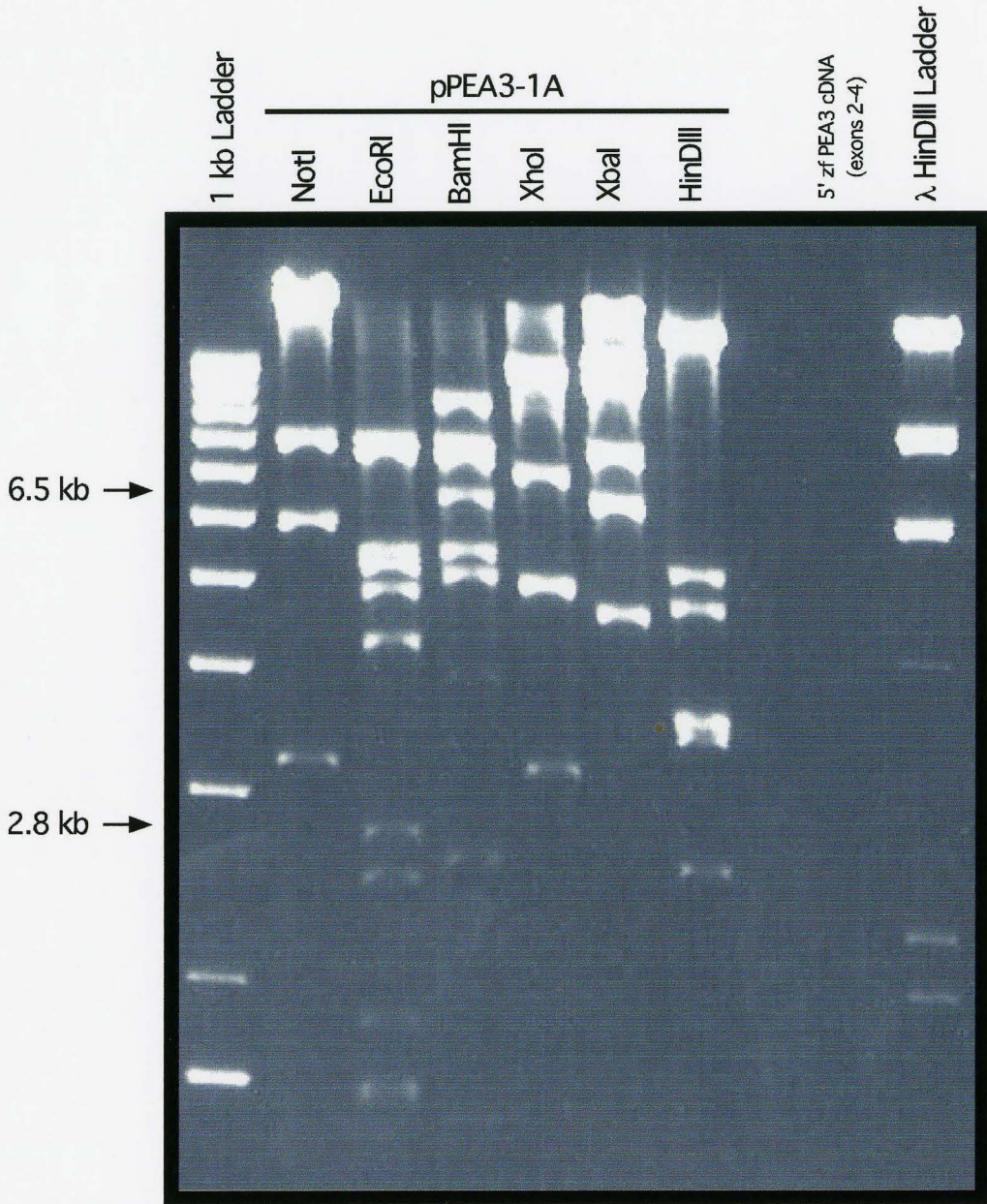
**Figure 13 Restriction Digestion of a Putative Positive Pufferfish Cosmid
from the Initial Library Screen**

The putative positive pufferfish PEA3 cosmid pPEA3-1 was digested with EcoRI, BamHI, HindIII, XbaI, and Sst I and the product digestion fragments separated by gel electrophoresis. The position of a 4.8 kb restriction fragment in the BamHI lane is indicated. A 1 kb and a λ HindIII DNA ladder have been included. As a positive control for future southern hybridization, 10 ng of zebrafish PEA3 cDNA (1.7 kb) was electrophoresed.



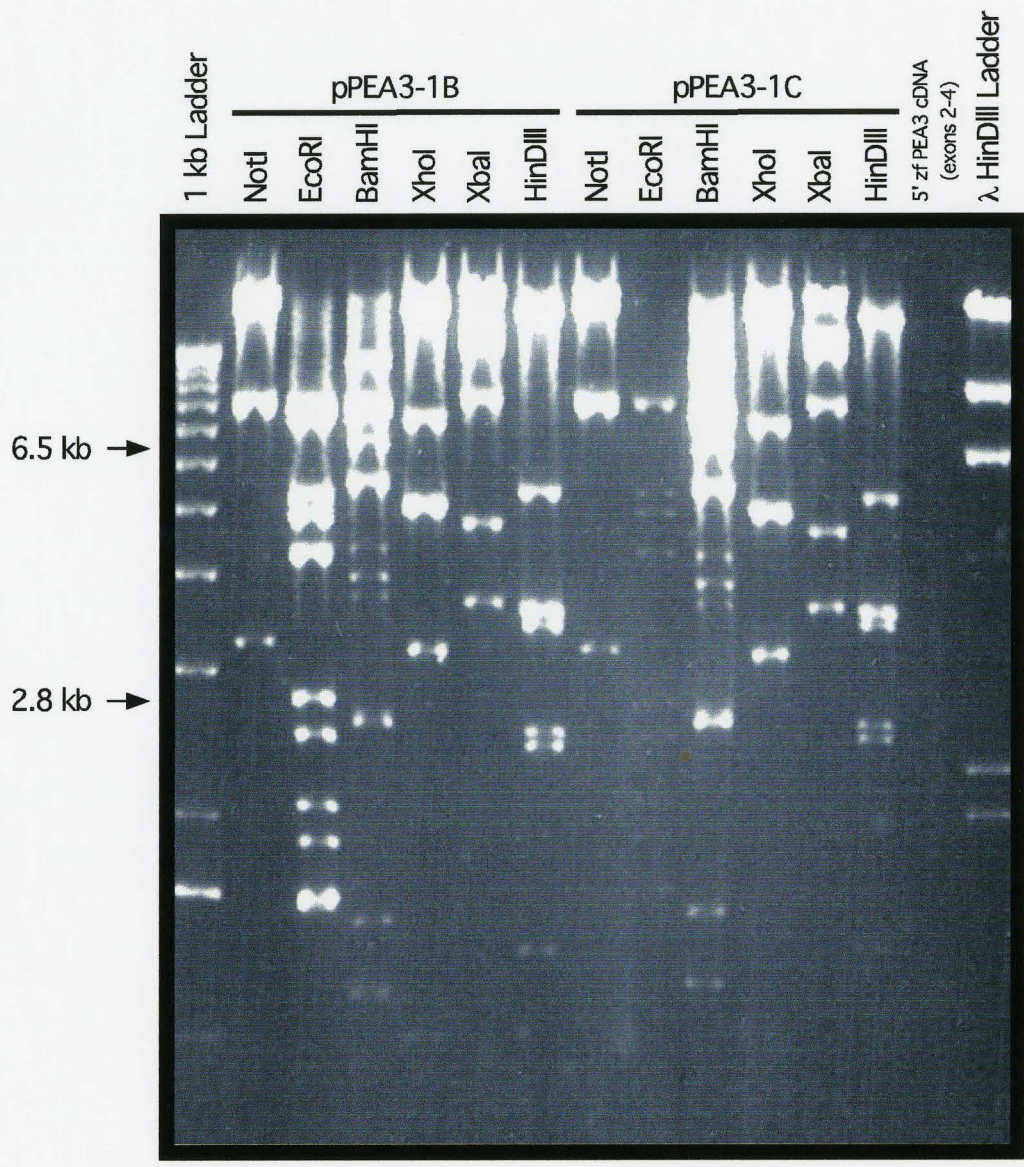
**Figure 14 Restriction Digestion of a Putative Positive Pufferfish Cosmid
from the Second Library Screen**

The putative positive pufferfish PEA3 cosmid pPEA3-1A, that was isolated by screening the genomic pufferfish cosmid library with the 5' end of the zebrafish PEA3 cDNA, was digested with NotI, EcoRI, BamHI, XhoI, XbaI and HinDIII and the product digestion fragments separated by gel electrophoresis. The position of a 6.5 kb restriction fragment in the BamHI lane is indicated. A 1 kb and a λ HinDIII ladder have been included. As a positive control for future southern hybridization, 10 ng of the 5' zebrafish PEA3 cDNA (215 bp), that was produced by PCR, was electrophoresed.



**Figure 15 Restriction Digestion of Two Further Putative Positive Pufferfish Cosmids
from the Second Library Screen**

The putative positive pufferfish PEA3 cosmids pPEA3-1B and pPEA3-1C, that were isolated by screening the genomic pufferfish cosmid library with the 5' end of the zebrafish PEA3 cDNA, were digested with NotI, EcoRI, BamHI, XhoI, XbaI and HindIII. The products of these restriction digestions were then separated by gel electrophoresis. The position of a 6.5 kb fragment in the BamHI lanes and the position of a 2.8 kb fragment in the EcoRI lanes are indicated. A 1 kb and a λ HindIII DNA ladder have been included. As a positive control for future southern hybridization, 10 ng of the 5' zebrafish PEA3 cDNA (215 bp), that was produced by PCR, was electrophoresed.



Examination of the EcoRI restriction digestion products from pPEA3-1A, 1B and 1C revealed that these cosmids had a number of restriction fragments in common. Specifically, the 1.5, 1.8, 2.6, 2.8, 4.2, 4.9, 5.2 and 8.0 kb restriction fragments appeared to be present in each of the EcoRI restriction products from these three cosmids. This finding suggests that these three cosmids share a considerable amount of sequence. The restriction fragment patterns produced for pPEA3-1B and pPEA3-1C were the same, indicating that these two cosmids are identical to each other.

Restriction digestion analysis was also performed on each of the twelve cosmids produced in the third library screen (Figure 16). Each cosmid was digested with EcoRI and the restriction fragments separated on a 0.8% agarose gel. Examination of the restriction fragment pattern produced by this digestion revealed that these twelve cosmids also shared restriction fragments in common with pPEA3-1A, pPEA3-1B and pPEA3-1C. Also a 1.5, 2.8, and 5.2 kb EcoRI restriction fragment appeared to be shared by all 15 cosmids. These results indicated that the cosmids produced by the screening of the genomic library with the 5' end of the zebrafish PEA3 cDNA had sequence in common with the cosmids derived from the use of the 636 bp pufferfish PEA3 genomic probe.

Based on the fact that a single XmaI restriction site exists slightly 5' of exon 5, a XmaI restriction digest was also performed on the 12 cosmids from the third library screen. This digest proved to be extremely useful in the identification of restriction fragments bearing genomic sequence surrounding exon 5 (data not shown).

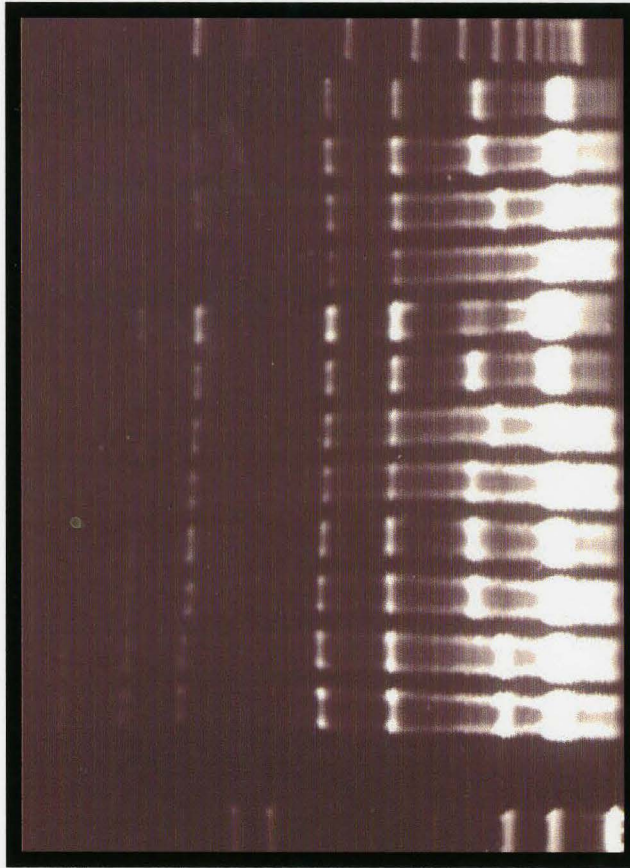
Southern Blot Analysis of Cosmid DNA

Following restriction endonuclease digestion of pPEA3-1 cosmid DNA, southern blot analysis was performed by hybridization of a full-length zebrafish PEA3 cDNA probe to DNA fragments that had previously been immobilized on a nylon membrane. The resultant autoradiogram revealed that a number of bands hybridized to the zebrafish PEA3 probe (Figure 17). Of particular interest was a single 4.8 kb

**Figure 16 Restriction Digestion of Putative Positive Pufferfish Cosmids
from the Third Library Screen**

The putative positive pufferfish PEA3 cosmids 019G13, 019G21, 037A22, 042K06, 073B12, 113F13, 117D03, 143G07, 184N11, 184M16, 196D02 and 196E02, that were isolated by screening a gridded genomic pufferfish cosmid library, were digested with EcoRI and the products separated by gel electrophoresis. The position of a 2.8 kb fragment common in each of the lanes is indicated. A 1 kb and a λ HindIII ladder have been included. As a positive control for future southern hybridization 25 ng of the pufferfish PEA3 5' RACE product in the pGem-T Easy vector was electrophoresed.

2.8 kb →



1 kb Ladder

019G13

019G21

037A22

042K06

073B12

113F13

117D03

143G07

184N11

184M16

196D02

196E02

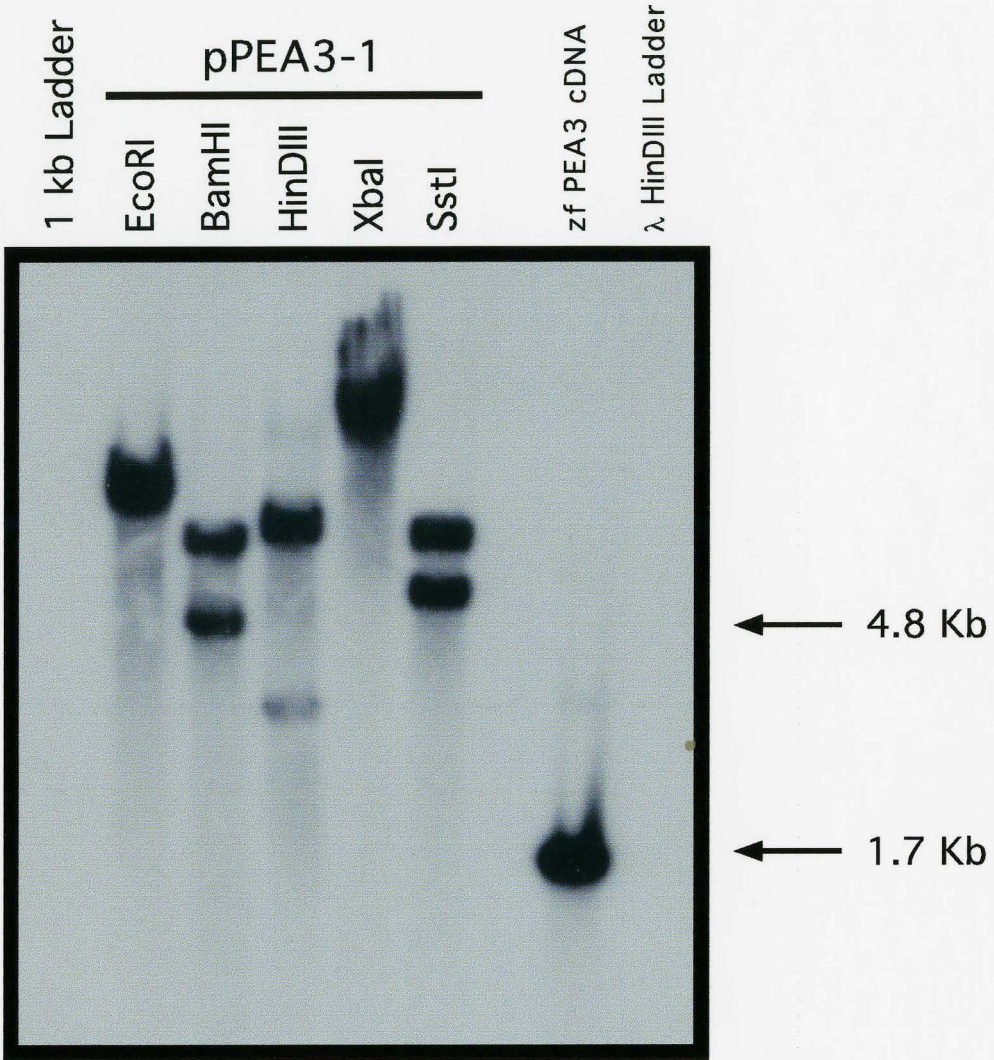
ECORI

25 ng Pufferfish PEA3
5' RACE in pGem-T Easy

λHindIII Ladder

Figure 17 Southern Blot Analysis of the Putative Positive Pufferfish Cosmid pPEA3-1

The full length zebrafish PEA3 cDNA probe was hybridized to the immobilized restriction fragments from the digestion of pPEA3-1 with EcoRI, BamHI, HindIII, XbaI and SstI. The size of the restriction fragment (4.8 kb) that hybridized to the probe in the BamHI lane is indicated. Further, the size of the zebrafish PEA3 cDNA positive control (1.7 kb) is also shown.



BamHI restriction fragment that migrated separately from the remaining BamHI restriction products. This fragment was used in subsequent subcloning experiments.

Southern blot analysis was also performed on the restriction fragments that were produced by digestion of pPEA3-1A, 1B and 1C with NotI, EcoRI, BamHI, XhoI, XbaI and HindIII. The probe used for this analysis was the same probe that had been used to isolate these cosmids during the second library screen (exon 2-4 of the zebrafish PEA3 cDNA). Examination of the autoradiograms from these southern blots showed that a 2.8 kb EcoRI restriction fragment hybridized to the 5' zebrafish PEA3 cDNA probe in each of the three cosmids under study (Figures 18 and 19).

A similar analysis was performed on the twelve cosmids that were isolated through the library screen using the pufferfish PEA3 genomic probe. In this case, the probe used for hybridization was the 334 bp 5' RACE product that contained pufferfish PEA3 cDNA sequence from slightly 5' of the putative start codon to exon 5. The autoradiogram from this southern analysis showed strong probe hybridization to a 2.8 kb EcoRI restriction fragment (Figure 20). Probe hybridization was also observed to occur to a band of an approximate size of 10 kb.

An additional southern analysis was also performed on the 12 cosmids generated from the screen of the gridded pufferfish genomic cosmid library. During this analysis immobilized XmaI restriction digest fragments were hybridized to the 636 bp genomic pufferfish PEA3 probe containing exon 5. This analysis revealed that the probe hybridized to a 900 bp and a 3.3 kb fragment (data not shown).

Subcloning of Pufferfish PEA3 Restriction Digest Fragments

From the southern analysis, a number of restriction digest fragments, which hybridized to various PEA3 probes, were subcloned into the pBluescript KS II cloning vector. The first fragment to be successfully subcloned was the 4.8 kb BamHI digestion product from pPEA3-1. This fragment was selected as a candidate for subcloning since it hybridized to the full-length zebrafish PEA3 cDNA probe.

Figure 18 Southern Blot Analysis of the Putative Pufferfish Cosmid pPEA3-1A

The partial 5' zebrafish PEA3 cDNA (215 bp) probe was hybridized to the immobilized fragments from the digestion of pPEA3-1A with NotI, EcoRI, BamHI, XhoI, XbaI and HindIII. The size of the restriction fragments (6.5 kb and 2.8 kb) that hybridized to the probe in the BamHI and EcoRI lanes are indicated. Further, the size of the partial 5' zebrafish PEA3 cDNA positive control (215 bp) is shown.

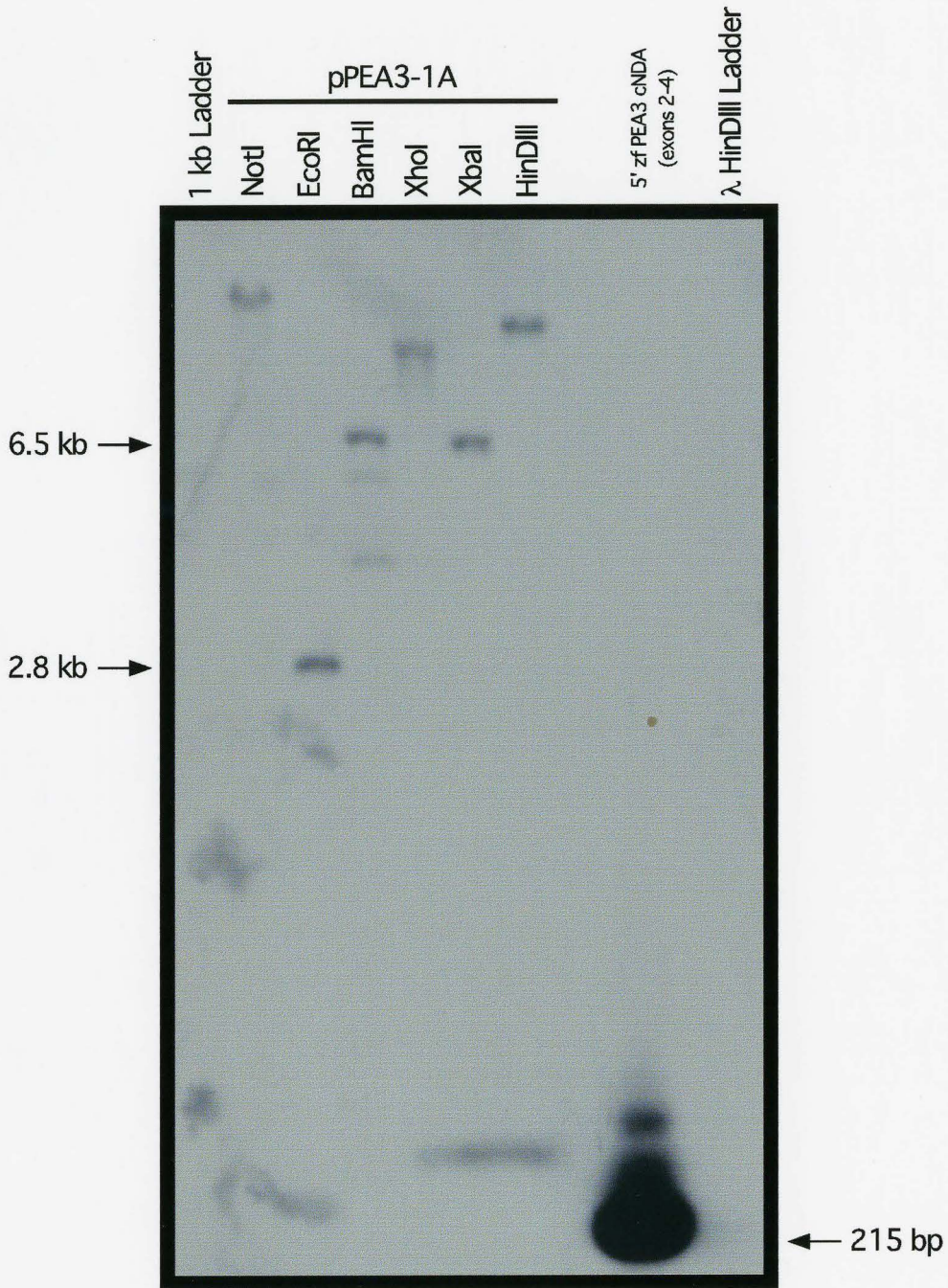


Figure 19 **Southern Blot Analysis of the Putative Pufferfish Cosmids
pPEA3-1B and pPEA3-1C**

The partial 5' zebrafish PEA3 cDNA (215 bp) probe was hybridized to the immobilized fragments from the digestion of pPEA3-1B and pPEA3-1C with NotI, EcoRI, BamHI, XhoI, XbaI and HindIII. The size of the restriction fragments (6.5 kb and 2.8 kb) that hybridized to the probe in the BamHI and EcoRI lanes are indicated. Further, the size of the partial 5' zebrafish PEA3 cDNA positive control (215 bp) is shown.

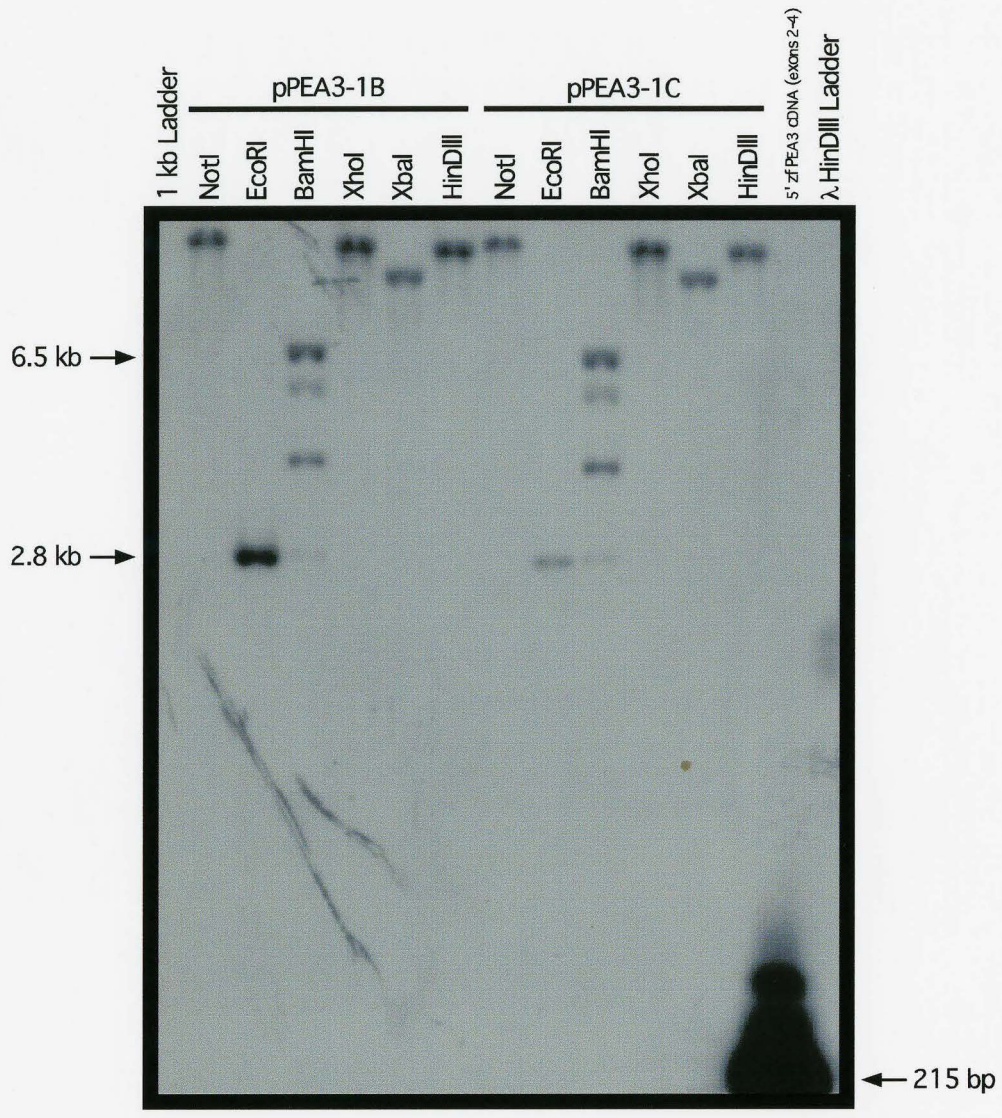
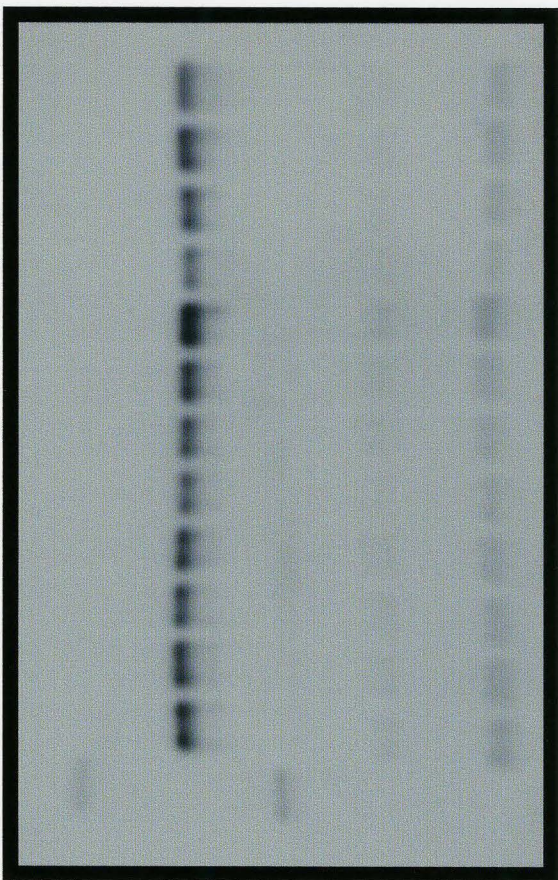


Figure 20 Southern Blot Analysis of the Putative Positive Pufferfish Cosmids from the Third Library Screen

The pufferfish PEA3 5' RACE product was used as a probe to hybridize to the immobilized EcoRI restriction fragments from cosmids 019G13, 019G21, 037A22, 042K06, 073B12, 113F13, 117D03, 143G07, 184N11, 184M16, 196D02 and 196E02. The size of the 2.8 kb restriction fragment that hybridized to the probe in each of the lanes is indicated.

2.8 kb →



1 kb Ladder

- 019G13
- 019G21
- 037A22
- 042K06
- 073B12
- 113F13
- 117D03
- 143G07
- 184N11
- 184M16
- 196D02
- 196E02

ECORI

25 ng Pufferfish PE3
5' RACE in pGem-T Easy

λHindIII Ladder

A second restriction fragment that was cloned was the 6.5 kb BamHI fragment that was produced by digestion of pPEA3-1A. Again, this fragment was shown to hybridize to the 5' zebrafish PEA3 cDNA probe (exons 2-4). The 2.8 kb EcoRI restriction fragment was subcloned from a wide variety of cosmids, namely pPEA3-1A, pPEA3-1B and 019G13. This 2.8 kb fragment also hybridized to both the 5' zebrafish PEA3 cDNA probe (in the pPEA3-1A, pPEA3-1B and pPEA3-1C southern blots) and the 5' RACE pufferfish PEA3 product (12 cosmids from the third screen).

The two XmaI fragments (900 bp and 3.3 kb) that had hybridized to the 636 bp genomic pufferfish PEA3 probe were also subcloned into the pBluescript cloning vector. The 900 bp XmaI fragment had been subcloned and forwarded to our laboratory, and the 3.3 kb XmaI fragment was cloned as a PspAI digestion product (a neoschizmer of XmaI) into pBluescript. A summary of the subcloned genomic fragments from pufferfish PEA3 is shown in Figure 21.

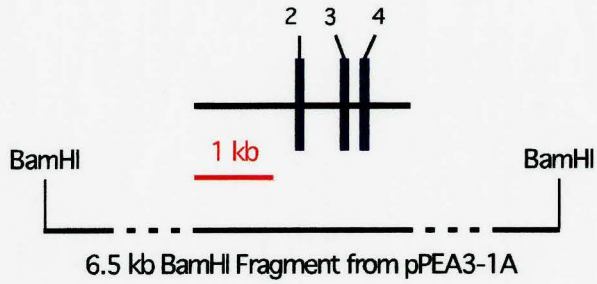
DNA Sequencing of Pufferfish PEA3

To this point the sequencing of genomic pufferfish PEA3 is incomplete. However, three continuous portions of the gene have been sequenced. The first, is 2.3 kb in size and contains the 5' end of the PEA3 gene including exons 2 to 4 as well as approximately 1 kb of sequence upstream of the putative start codon. This sequence was derived through sequencing of the 6.5 kb BamHI fragment from pPEA3-1A using primers specific to exon sequence from the 5' RACE product. The second genomic fragment of pufferfish PEA3 that has been completed encompasses slightly over 4.0 kb of sequence. Within this region only exon 5 has been identified based on sequence comparison to the 5' cDNA sequence that was determined through 5' RACE analysis. Exon 5 was found to reside approximately 900 bp into this genomic sequence. Analysis of the remaining 3.1 kb of downstream sequence, via alignment to other PEA3 orthologue cDNAs, including that of zebrafish PEA3, failed to allow for the identification of further potential exons. Since this approach failed to clearly denote the position of further exons, it is reasonable

Figure 21 A Summary of Subcloned Restriction Fragments from the Pufferfish PEA3 Cosmids

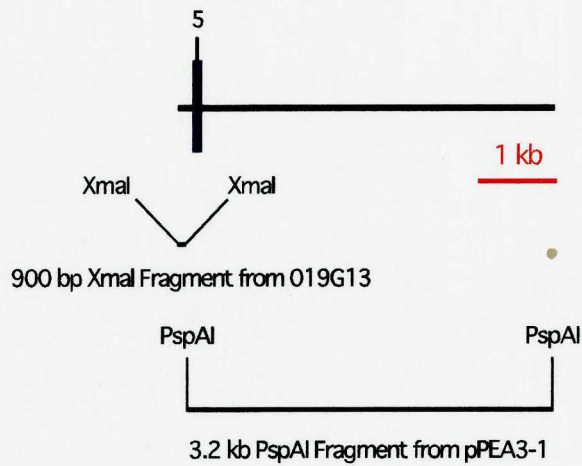
- A. The intron-exon organization of the 5' end of the genomic pufferfish PEA3 gene is depicted. Exons have been indicated as blue boxes with the corresponding number given above. In this case, exons 2, 3 and 4 have been shown. A 1 kb marker is also present to indicate the scale of the illustration. Below, is a summary of the restriction fragments that have been cloned which contain exons 2, 3 and 4 as well as a region of non-coding sequence 5' of exon 2. The restriction sites are indicated at either end of the fragment and the parent cosmid from which the fragment was subcloned is indicated. Broken lines indicate that this region of the fragment has not been sequenced.
- B. The central portion of genomic pufferfish PEA3 is presented. Exon 5 has been depicted as a blue box with the exon number given above. A 1 kb marker is also present to indicate the scale of the illustration. Below, is a summary of the restriction fragments that have been cloned which contain genomic sequence within this region. The restriction sites are indicated at either end of the fragment and the parent cosmid from which the fragment was subcloned is indicated.
- C. The intron-exon organization of the 3' end of the genomic pufferfish PEA3 gene is depicted. Exons have been indicated as blue boxes with the corresponding number given above. Here, exons 12 and 13 of pufferfish PEA3 are shown. Further, the final two exons of the pufferfish homologue of human RNA helicase 1 (HRH1) are shown. A 1 kb marker is also present to indicate the scale of the illustration. Below, the single BamHI fragment, from which the sequence was generated, is indicated.

A.

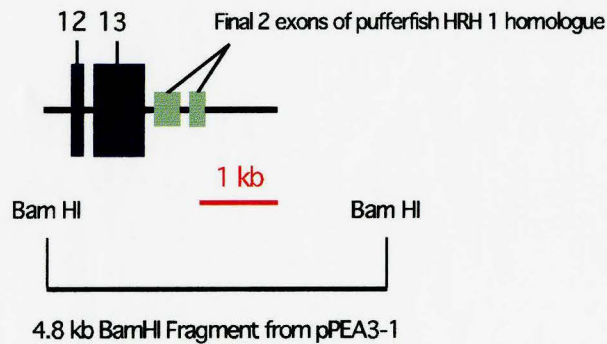


2.8 kb EcoRI Fragment from pPEA3-1A, pPEA3-1B, pPEA3-1C and the 12 Cosmids from the Third Library Screen

B.



C.



to assume that the sequences of exons 6 to 11 in pufferfish PEA3 are highly dissimilar to any other orthologous PEA3 sequence. As such, the central and 3' end of the pufferfish PEA3 cDNA will be required to clearly identify these exons within the genomic sequence. This sequence was produced by direct sequencing using cosmid pPEA3-1 as a template, coupled with the sequencing of the 900 bp XmaI fragment from cosmid 019G13 and the 3.2 kb PspA1 fragment from cosmid pPEA3-1. The third continuous portion of genomic sequence that has been produced encompasses just under 2.5 kb of sequence and resides at the 3' end of the pufferfish PEA3 gene. This sequence was found to contain exons 12 and 13 as well as the pufferfish homologue of the human RNA helicase 1 (HRH1) gene. Genomic sequence in this region of the PEA3 gene was produced by sequencing a 4.8 kb BamHI fragment from the pPEA3-1 cosmid. The partial genomic sequence of pufferfish PEA3 is presented in Appendices C, D and E.

Four-way Alignment of the Mouse, Human, Chicken and Pufferfish PEA3 Promoters

Sequence Identity Between the Promoters

Alignment of the mouse, human, chicken and pufferfish PEA3 promoters revealed that directly upstream of the transcriptional start site (as identified in the mouse and human sequence) a significant amount of four-way sequence identity could be found. Specifically, this region was found between positions -260 and +1, relative to the mouse PEA3 promoter. As well, a striking 14 bp region of four-way identity was observed between positions -254 and -240. Further sporadic sequence identity may also be observed to occur elsewhere within the alignment. The results of the four-way alignment of the promoters is shown in Figure 22. Two-way alignment of the pufferfish PEA3 promoter with each of the mouse, human and chicken PEA3 promoter sequences may be seen in Appendices G through I. These two-way alignments illustrate the sequence identity that is observed between each sequence pair.

Inspection of the four-way promoter alignment also reveals that in the case of the mouse, human and chicken genomic sequence, the putative start sites of transcription, the putative start sites of translation and exons 3 and 4 are well aligned with one another. However, when the pufferfish genomic sequence is aligned, exons 3 and 4 are found to perfectly coincide with the other sequences, but the putative start codon in exon 2 of the pufferfish gene aligns upstream of the start codons in the other three genes.

In addition to the four-way sequence identity that was observed, a much higher amount of three-way sequence identity may be found between the mouse, human and chicken PEA3 promoters. The region of three-way identity may be found to span from approximately position -300 to the transcriptional start point at +1.

Identification of Conserved Consensus Transcription Factor Binding Sites Within the Promoters

Within the region of four-way promoter alignment from -260 to the transcriptional start site at +1 a number of different conserved consensus transcription factor binding sites are present. Those binding sites that are only conserved across three of the promoters include: 2 HNF3 β , 2 NFY, 1 SRY, 1 TCF, 2 AP-2, and 1 v-myb. Outside of the region of sequence conservation a single AP-2 and a single c-Ets-1 binding site may also be found within these three promoters. A limited number of transcription factor binding sites are also conserved in each of the four promoters. These binding sites include: 1 SRY, 1 AP-1 and 1 σ EF1. The positions of each of these consensus binding sites may be seen in Figure 22.

Common Features within the PEA3 Promoters

A number of other features also become evident through the production of the three-way alignment of the mouse, human and chicken PEA3 promoters. Firstly, at position -22 just 5' of the transcriptional start site, there appears to be a conserved CCAAT box common to all three promoters. This sequence matches very closely (10 positions out of 12) with the recognized CCAAT consensus sequence (Bucher, 1990). Secondly, surrounding the major start site of transcription, the initiator or cap signal may be readily observed in the mouse and human sequence. The alignment also shows that in the chicken the initiator element is also present and very closely matches with the consensus (6 positions out of 8) (Bucher, 1990). Lastly, both the mouse and human cDNA and genomic PEA3 sequences contain two putative start sites of translation within exon 2. Inspection of the genomic chicken PEA3 sequence reveals that it only contains one putative ATG codon, that through alignment, matches most closely with the second start codon in human and mouse PEA3. Further, the first start codon in the human and mouse sequence does not lie within an optimal Kozak consensus element. The second translational start site however does possess an adenosine residue at position -3, and this position is conserved across all three sequences. This is of interest since an adenosine residue at position -3 has been shown to be the minimal requirement to support translational initiation (Kozak, 1991).

Figure 22 **Four-way Alignment of the Mouse, Human, Chicken and Pufferfish PEA3 Promoters**

The existing sequence upstream of exon 4 of mouse, human, chicken and pufferfish PEA3 was aligned using the Clustal Alignment subprogram of the DNASTar software package. The alignment has been numbered in relation to the mouse PEA3 sequence with +1 indicating the start site of transcription. Four-way positional identity has been indicated in green throughout the alignment. The defined exon sequences from mouse, human and pufferfish PEA3 are boxed and indicated at the beginning of each exon. No exons have been identified in the chicken sequence because a 5' chicken PEA3 cDNA has not been identified. Putative consensus transcription factor binding sites that are found to be conserved in three or four of the promoters, and which are 100% identical to a core transcription factor consensus binding site and 85% identical or greater to the adjacent sequence within the consensus site have been included. The names of the transcription factors have also been indicated. The putative start codons of mouse, human, chicken and pufferfish PEA3 have been shown. Further, the -3 position upstream of the second in-frame start codon in the mouse, human and chicken sequence is circled in red. The position of a putative CCAAT box and a consensus initiator or cap signal is shown.

Mouse PEA3 Promoter	G A - A T T C T G G G T C A G A C T G A G T T G A G A G A C C C T G T C T C A A A A A C A T G G T A
Human PEA3 Promoter	G G - G - - - A G G T T G A G G T G - - T T A G G A - - - - T A G C T T G A A G C C A G G G - -
Chicken PEA3 Promoter	A A - A - G G A C T T G G A C T C - - - C C A T G A C C
Pufferfish PEA3 Promoter	G G T A C T T T G G A T C G G A C C G - - - - - - - - - C C T - - - C T G G A G C T - - - - -

- 3 kb

Mouse PEA3 Promoter	A G G T A A G T A G G A A G G A A G G A A G G A A G G A A G G A A G G A A G G A A G G A A G G A A G
Human PEA3 Promoter	- - - - A G T T T G A G G T T A T - A G T A A G C T A T
Chicken PEA3 Promoter	A G - A A G A G C A A A A G C A A - - A A G G C T G C A G G - - - -
Pufferfish PEA3 Promoter	- -

Mouse PEA3 Promoter	G A A G G A A G C C A C A A C A A A C C A T G A G C T G A G A G A C G G T G C A T C A G T C T A G A
Human PEA3 Promoter	G A T G G C A N C C A C T G T A C T C C A - - A C C T G G G T G A C A G A G - - - - - - - - - G G A
Chicken PEA3 Promoter	- - - - - A A C A G C A G C A A T C C A C - - - - - A G A G C T G G A A T G T T C A T - - - G A
Pufferfish PEA3 Promoter	- - - - G C A G C C - T C T G G A

Mouse PEA3 Promoter	G C A C T T G C T G C T T T T T C A G A G G A C C T G A G T T C T G T T C C C A G C A C C T A C A T
Human PEA3 Promoter	G A C C C T C C T G T C T C T T A A A A A G A A A A A A A A A - - - - A A A A A A A A A A
Chicken PEA3 Promoter	G - - - - - G C T G A T T T A G A A T G - A C A C
Pufferfish PEA3 Promoter	G - - - T T G C A G C C - - T C T G A G -

Mouse PEA3 Promoter	C A G A C G G C T C A C G A C T G C C T G T T A C C T G C C G C A T C A G G A A A T C C A G T G T C
Human PEA3 Promoter	A A A A A G G C T - - G G - - G C T T G G T G G C T - - C A T G C C T G T A A T C C C A G C A C T
Chicken PEA3 Promoter	A A A A A C A C T C A A A A G T G T T - - - - - - - - - - - C C T G G G - - - T C A G G A T C
Pufferfish PEA3 Promoter	- C T G C A G C C T C T G G A G C T T T A G A T C C

Mouse PEA3 Promoter	T C T G G C C T C T G T A G G T A C C C A C A C A A A C C T A C A C A A A C A C A T A A A T A A A A
Human PEA3 Promoter	T - - - - - - - - - T A G G A A G C C A A G T G G A T C - A C C T G A G G T C G G G A G T - - - -
Chicken PEA3 Promoter	T G C A T T C T T T - C A G A T A T T C - - - - - - - - - C A T C A A C T T A A G G A T T G A A G
Pufferfish PEA3 Promoter	T G C G G T T T C C G T G T G - - - - - G C G C A G A C C T A - - - - - - - - - - - A A T A A A A

Mouse PEA3 Promoter C T C C A T T G G A G A A G A G G T C C A A T C T C C C T T G G A G A A A T G T C A T T G A T C A T
 Human PEA3 Promoter C T T T A T T G T A - - - - - C C T A T T C A C T A C A C A A G C A T T T C A T - - G C C - -
 Chicken PEA3 Promoter - C C C A C T G G A T G G G A A A T T - - - - - - - - - - - - - - - A G A A C A T C A T A A A G C A -
 Pufferfish PEA3 Promoter - - - - - - - - - G A G A G A - - - - - - - - - - - - - - - A A A A C A T G T C G - - - - - - - - -

Mouse PEA3 Promoter T G A T C A T C A T C T G A G T G T T C A G T C C C A G G C C A G G G C A G G T G T G G C A G C T G
 Human PEA3 Promoter - - A C C C T C A T A A G C A T G C T A A - - - - - C A G G C T G A G G - A G G - - - - - - - A C T T
 Chicken PEA3 Promoter - A A C A A A T A T A A C A G T G G A A A A T C C C A T - C C G G T T C A G - - - - - - - - - - -
 Pufferfish PEA3 Promoter - - - - - - - - - G G A G T - - - - - - T T T C G T G C C A G G G C A G G T T T - - - - - T T A

Mouse PEA3 Promoter T A G C T G C A G C T G C A G T T G G T T C T C T T G C T G G G T C T A T A G C T T G A G C C A G A
 Human PEA3 Promoter T - - C A G A G G C C A - A A T G A G C T C T - - - - - G G - - - A G A G C T C T - - C C A T A
 Chicken PEA3 Promoter - A G C - G C A G C C A T A - - - - - - - - - - - T G C T G A A A T C A G A T C T - - - - - - - G A
 Pufferfish PEA3 Promoter T G A A T G G A G C C C T - - - - - - T T T C T T T C T - - - - - - - - - C T C - - - - - - - - -

Mouse PEA3 Promoter G A T A G G G A G G T A G G C T T C T T T G G T C T C T G C C A G A C C T G G G A A G C A G A G G A
 Human PEA3 Promoter G A T A A G G A T G T C T A T A C C A T G A G G C A G G A T G A A A A C A T C C A - C A C C A G A
 Chicken PEA3 Promoter A A T A A C A A A G - - - G T T C C A T T A - - - - - T A T C T T A A C C A G G A A A T G C - - - -
 Pufferfish PEA3 Promoter - - - - - - - - - - - - - - - T C T C T C T -

Mouse PEA3 Promoter A A T A T T A A C T A G C C A G G G C T C T C C A G A G A A G T G G A G G A A C T C C G A G T C A G
 Human PEA3 Promoter G A T A T - A G T C A - - - A G A A T G T T G T A T A T G G C T A A A G G A A C T - - G A T T C A G
 Chicken PEA3 Promoter A G C A T T A A T A A C C A A G A G T - - - - - T A A A T G G C T - - - - - A T T T T G - - - - -
 Pufferfish PEA3 Promoter - - - - - - - - - - - - - - - T T T T G T G A A T - - - - - - - - - G A A - - - - - - - - - A A

Mouse PEA3 Promoter A C A G G C A G C C C C G G A C A G G C C T C A A C T G G A G C A C C C A G T G C T C A C T C A G G
 Human PEA3 Promoter G C T A T C A - - - - - T G G A - A G G C T T C - - C T G G A G G A - - - G A G G G A C A A T - A G -
 Chicken PEA3 Promoter - C A A T - - G A C C T G T A T A G - - - - - - - - - C A G A A A G T G C A A A G C A - A A G C A G G
 Pufferfish PEA3 Promoter A C A A - - - - - - C T G G A A A A G C C A C A A T - - - - - A A A A C C C A G - - - - - - - - - C A G -

Mouse PEA3 Promoter G A C C C A C C T G A A C C T T C A T G G A G A C C A C A T C A G A G A T C T T T T C C T A G A A C
Human PEA3 Promoter - - - - A C A T G A G C C T T - A T A G A - - - - A T A T T A C A G C A C C T A T - - - A G A A A
Chicken PEA3 Promoter G G - - T A T C T G G - - T G T C G C A T A G A C T - - - T T G C A G G A C A - - - C T T A A A - -
Pufferfish PEA3 Promoter - - - C T G A T T G G - C T G C C G C G C A A - - - - - - - - - - - - - - - - C C T C G G - - -

Mouse PEA3 Promoter T G C C T G C C T T C C T C C C C A C T T C T C C T G G T C T A G G A A C T C A C A C C C C T C C
Human PEA3 Promoter G G G A A G C T T A C T T C - - - C A T T T - - - - - G C A C A A G - - - - C A G A T C T C A A T
Chicken PEA3 Promoter - G T A T G - - - - - - - - - - - - - - - A A T T - - - - - G C T T A A C A G C - - - A A T C T - - - -
Pufferfish PEA3 Promoter - - - - - - - - - - C C C C G C C C A - - - - - C C G A C - - - - - G T T - - - A T C - - - - -

Mouse PEA3 Promoter C T A A A C T G G C T C C A T C C G C C C C A A G C C G G G A T T T G C T T G G G C T T G C C T G T
Human PEA3 Promoter G G A A T G A G C C C T T A T G A A C A C T A A A C - A G C A A T T G T A T G T T T T A G C C C T - T
Chicken PEA3 Promoter - A A C T G G G G G A C G T A T - - T C T - - C T T G C
Pufferfish PEA3 Promoter - - - - - - - G T C C T C A T - T C C A G C - - - - -

Mouse PEA3 Promoter C T T T C C T T G G G C A G G - G T G C C T G C G G G G C T C C G T T G C T T A G C T T T T T A T T
Human PEA3 Promoter A G T T C - - T G G C C A G A C A G G C C C A A A G A G C T T A C A T G T G T C A T T T C C T T C T
Chicken PEA3 Promoter A G C T T C C C A G - T A G C A T T C G A T T
Pufferfish PEA3 Promoter - A C T T C C G A T T

Mouse PEA3 Promoter G T A G C T G T T C A C C T G A C A G G C A T T T T G T G C A A C T C T T G A G T G C G A A G C G A
Human PEA3 Promoter C T G A A A C T C C A C - T G C C T - - C A T C T - G T A - A A G T G G C G G G T G T T A G G T T A
Chicken PEA3 Promoter G A A G - - - - T T A A C T G - - - - - - - - - - - C T T G A A C A A A T T G C - - - T A C A A A A C A A
Pufferfish PEA3 Promoter -

Mouse PEA3 Promoter G G C T C A A G G A A G G A G T C - C T C C T C T G G G C C C C G C A G A G A T C C C C A T A A C
Human PEA3 Promoter G A T G A A T T T G G G A A G C C A C T C C T G C T G G C C T - - - - - G G C A T T T G A A C A T T
Chicken PEA3 Promoter T A T - C T G T A T A T A T C
Pufferfish PEA3 Promoter - G G C C C G C G T G G - - - - - - - - - - - - - - -

Mouse PEA3 Promoter T A T A T A G G G A T T T C T C T A A C A G G G A A G A G A A T G A A A G A T A C C T C A C C A T A
 Human PEA3 Promoter C T T T T G T G T T T A T C A C A A A C A G G C A G G A - - - T A A A A A A T - - - - -
 Chicken PEA3 Promoter T G T A T A T A - A T A T C C C C A G T A - - - - - A T A T C T G G G G A T G C - - - G C T A T A
 Pufferfish PEA3 Promoter - G A G A G G A G - - - G G G G G G T T C A T - - T C A T A

Mouse PEA3 Promoter T A T A T G G A C T G A C C C A G G C T C T C A T G A A A A G G G G G T C C C T A G A G G A G A G
 Human PEA3 Promoter - - - - - - - - - - - C A G G C T G T C C T C A A A - - - G G A G T G C A T A G G C T G G A G
 Chicken PEA3 Promoter A G T - T A G C A T G T T C - - - - - T T G T T T T - - - - - - - - - G C C C C C A - - - - -
 Pufferfish PEA3 Promoter A G T G - A G C G T A T T C - - - - - - - - - - - A A G G G G A - - - - - C A A G G G G A G

Mouse PEA3 Promoter C G G C C A G A G G T T C A A A C T T T A A A G A A C A G T T T G G T A G G T A A A G G G A G G A
 Human PEA3 Promoter - A C A T A G A G A A T T G - - T C T C A G C A A A A A G - - - - - G C C T T T G A G G A T T C A T G
 Chicken PEA3 Promoter T G A A C A G A - - - T T A A A C C T - - - - - - - - - - - G C T T T T G G G G - - - - -
 Pufferfish PEA3 Promoter - - - - - - A A A T T C G A G C T C T T G A A A - - - - - - - - - - - G A G G A - - - - -

Mouse PEA3 Promoter G T T T G T G T G G C A G A A A G A G C C C A G T T C A G G T G C C A G G G G T T A A G G T T T T
 Human PEA3 Promoter T C C T T T G T G C C C C C A T G G C C A C A G C C C A A G C A C C - - - - - C C A G G T - - -
 Chicken PEA3 Promoter - - C T G T T T - - - - - - - - - - - - - - - T C C T G A C A - - - - - G A G T - - G G T T C C
 Pufferfish PEA3 Promoter - - - - - - - - - - - - - - - - - A A T C C - - - - A C C A A - - - - - T C A G G - - -

Mouse PEA3 Promoter T G T C C C T T C T C T G A C T C C T A C A G G C T G G A A G A G A T T G C A T G C A C C T T T T C
 Human PEA3 Promoter - - - - - C C C C T C A G A T G T C T G T - - - C T T G A C A T A G C T G C T T G C - - - - - C T G
 Chicken PEA3 Promoter T G G T - - - - - A A T G C T - - - - - - - - - - - - - - - G A A G C T G C - - - - - C T T T T T
 Pufferfish PEA3 Promoter - A G G C G A A - - - - - - - - - G C A A C T T T - -

Mouse PEA3 Promoter T T T C T C A G A A G T A C T G T T T G G T T A G A C A G A T G A A C T T T G A A A G C A G C C C T
 Human PEA3 Promoter A A A C C C A G C A C T G - T A T C T G A C A C G A T A A A T G T T T A T T G A A A G A A G T G G T
 Chicken PEA3 Promoter - T T T C C A G C A - - - - - - - - - - - - - - - G A A A G A T - - - - - T A A A A - - - - -
 Pufferfish PEA3 Promoter - A T C C C C G A A T - - C T G T C C G G C - - - - - A G G C G - - - - - - - - - - - - - -

Mouse PEA3 Promoter G C C A G G A G C C A G C C C C T A C T T T C T G A T T G G T C T C T C G T G G C T G G G A A A C T C
 Human PEA3 Promoter G C C G G G A G C C A A C C C C T A C A T T C T G A T T G G C C C C G C G G C T G G G A A A C T C
 Chicken PEA3 Promoter C A C C A G T G C - - G C C - - - G C G C T C T G A T T G G C C C C G C G G G A T - - A A A G C C A
 Pufferfish PEA3 Promoter - - - - - C A G C C T C C A C G C - - - - - A G C T C C A T T G A C G C G C A G C T C

- 156 NFY

Mouse PEA3 Promoter C T C C C T C A A G C T C A G C T A G C C G A G T C T G A G G A T T G - G C G G G C G T G C C C G G
 Human PEA3 Promoter C G C C C C C G G C C A G G C C G G A C G A C T C T C C G G A T T G - G C G G G C G T G C C C A G
 Chicken PEA3 Promoter C G C C C A C T T T C C C - - C C C A G C G T G - C T G C G - A T T G - G T G G A C G C G C C C A
 Pufferfish PEA3 Promoter C C T C C T - - - - C T C A T C T G A - - G T C C C T G C G A G C T G C G C G G A C C A G C T C A G

NFY

Mouse PEA3 Promoter C C C T G - - - G G G T T T C A T T C A C A A A A G T C A A T T G A A A C A A A A G G G A A C C G G T
 Human PEA3 Promoter C C C T G C T G G G G T T C C A T T C A C A A A A G T C A A T T G A A A C A A A A G G A A A C G G G G
 Chicken PEA3 Promoter G C C C - - - A G G G C T T C A T T C A C G A A - G T C A A T T G A A A C A A A G - - - - -
 Pufferfish PEA3 Promoter C C C C G A T G A - - - - - C C A C A A A - - - C G G T T A A - - - - -

AP-1 AP-1 HNF3β TCF SRY v-myb

1st Nucleotide of 5' RACE Product (Exon 2)

Human PEA3 Footprint -7

Mouse PEA3 Promoter G G - G G A G G G G C G G G G - G A A G A G C G - A G A A C G G A G G C C A A G G C A A A G G A A A
 Human PEA3 Promoter G G C G G T G T G G C G G G G C G G G G G T G G A G A G C T G A G G C C G A G - C G A A G G A A A
 Chicken PEA3 Promoter - - - - - A G C G G G A G C C G A G C T C G T G A A G G C G G C G A
 Pufferfish PEA3 Promoter - - C G G - - - - A T G G A T T A T A A G A T G G A - - - - - T G G A T A

Pufferfish PEA3 Start Codon -1 0 +1 +4 2nd in-frame Start Codon Exon 1 (+1) Initiator / Cap Signal

Mouse PEA3 Promoter T G C A C C A A T C A G C T G C T - C C C C C G G G C T C A C A A C T G T C T G C T G C G C C C G G
 Human PEA3 Promoter T G C A C C A A T C A G C T G C T - C C C C C G G G C T C A C A A C T G T C T G C T G C G C C C G A
 Chicken PEA3 Promoter T G C G C C A A T G G G A G G C C G C C C C G G G C G C T C A A C G G C T G - - - - -
 Pufferfish PEA3 Promoter T - - - - - T T G G A C C A G C A A G T G C C T - - T A C A C T T T -

+ 21 Putative CCAAT Box AP-2 AP-2 v-myb

Mouse PEA3 Promoter A A A A C A A G C C G G G G C T C T G G G G A C C C G G G G C T C A G G C C G C C T C G T C C G G
 Human PEA3 Promoter A A A A C A A G T C G G T A C G N T G G G G A C C C G G G G C C G G G G C C G C C T T A T T C C G
 Chicken PEA3 Promoter - - A A C G - G C C G - - - C G C - G G G - - - - C G G G G C T - - - - -
 Pufferfish PEA3 Promoter - - A G C A A A T - - - G T A C G - T A T G G A T G C A G - - - C T C - G A C C G - - - - T C C G G

c-Ets-1 (GGA Core)

Mouse PEA3 Promoter	A T C T C C C G G A A A T G G G A G C T T G G G C G A A G C G C T G A T G G T C C C G C A G G G A A
Human PEA3 Promoter	A T C G C C C G G A A A T G G G A G C T T G C G C G A A G C G C T G A T C G G C C C G C T G G G G A
Chicken PEA3 Promoter	- - - - C C G G A - - - - - - - C T C C G A C G G - G C G C T G - - - - - - - - C A A G A
Pufferfish PEA3 Promoter	A T C G C A T G G A A A T G G G C C C T A A A C A G A C T G T T A A T G G C A G C C A G A G G A

		δEF1
Mouse PEA3 Promoter	A G C T C A T G G A C C C G G G C T C C C T G C C G C C T T C C G A C T C A G A A G G T A A A G -	
Human PEA3 Promoter	A G C T C A T G G A C C C G G - C T C C C T G C C G C C C C T C G A C T C T G A A G G T A A A G A G	
Chicken PEA3 Promoter	G C C C C A T G G A G C C G G G C T T G C T G C C G G C G C A G G A T T C G G A A G G T A A - G G G	
Pufferfish PEA3 Promoter	A A T A C A T G G A A A C A G A - - - G C T G C C A C C T C A G G A A T C C G A A G G T A A A A - -	

Mouse PEA3 Promoter	- - A A A C A G A C A A A C T G C G A G N C - - C T C T G C T T A A T C C A C G T A C A T C C A T C
Human PEA3 Promoter	C A A G A G A G A G A A A C C G C G A C C C T C C T C C G C C A C A C A C A C A C A C A C A C A
Chicken PEA3 Promoter	T G G G G - - G G G G G G C T G T A G G G -
Pufferfish PEA3 Promoter	- -

Mouse PEA3 Promoter	C - - - - C A A A G C C T T A T C C T G G A C - - - T T C C C C A A C C C A C G - - - - T T T C C
Human PEA3 Promoter	C A C A C A C A C A C A C A C A C A C A C G G C C G A T T C T G G A A C C C C G C C T G T T T C C
Chicken PEA3 Promoter	- G A G G C T G G G G A A G G C T C T G G G C A G G G T C A G G T C C T C C C C C T G T G C C T
Pufferfish PEA3 Promoter	- - - - - C A A A -

Mouse PEA3 Promoter	C C A A A A G - - - - - - - - - T C A A A - - - - - C C A C T C A T T C T T C T C C G T C C C T T A
Human PEA3 Promoter	C C A A A A A A C T C G C C C T C C A A G G G A T T C T G G T C A T T C T T C T C G G T C C C T - A
Chicken PEA3 Promoter	C C G - - - - - - - - - C C C C A - - - - - C T C C C C G T C C C G C C C G G C A T C T C C
Pufferfish PEA3 Promoter	C - - - A A T A C T T G A C C T C - - - - - G G C T C A G G T T A T G A T - - - - - G A A C C T C -

Mouse PEA3 Promoter	T C A C C A G T G G G G A G G C T G T A A C C C G G G C T C G C C A G G - T C T C T T C C C - C A T
Human PEA3 Promoter	C T C C C A A G G A G G G A - - G T C A C C G G G T C T T G C T A G G G C C C C T T C T C T C T T
Chicken PEA3 Promoter	C C G C T - - - - - - - - - G G A G C - - - - - - - - - A T C C C G C T - - - G T C C C T C C C C - - - -
Pufferfish PEA3 Promoter	- - - - - - - - - - - - - - - G G G A C - G T G T C T G A G T C - - - - - G G A C C C A G C T T C T C T A

Isolation and Characterization of Pufferfish ERM Bacteriophage λ Clones

Library Screening and PCR Generation of a Pufferfish ERM Specific Genomic Probe

The initial screening of the genomic pufferfish λ bacteriophage library was conducted in the attempt to isolate phage clones whose inserts bore pufferfish PEA3 sequence. To this end, full-length mouse PEA3 cDNA was used as a probe in the screen. This first attempt at isolating clones containing pufferfish PEA3 sequence produced ten bacteriophage λ DNA species, however, initial sequencing of these clones revealed that the genomic sequence of the related Ets gene ERM had been discovered. Based on this finding, the ten clones from the first screen were given the designations: pERM-A, pERM-B, pERM-C, pERM-D, pERM-E, pERM-F, pERM-G, pERM-H, pERM-I and pERM-J. Following the initial isolation of pERM-A, both subcloning and sequencing from this clone allowed for the generation of genomic sequence from slightly 5' of exon 4 to exon 13 inclusive. Once it was determined that pERM-A did not contain a large amount of pufferfish ERM promoter sequence, a second library screen was started in an attempt to isolate further phage clones containing distal ERM promoter sequence. The genomic probe for this screen was produced by the design of intron specific primers on either side of exon 4 (AB. 12422 and AB. 12423), followed by the use of these primers in a PCR experiment. The use of pERM-A as a template for PCR produced a single product of 450 bp (Figure 23). Using this genomic probe to screen the pufferfish λ genomic library produced four new phage clones: pERM-2A, pERM-2B, pERM-2C and pERM-2D.

Restriction Digestion of Putative Positive λ Bacteriophage Inserts

Putative positive λ phage DNA from the first library screen was subjected to restriction endonuclease digestion using XhoI (Figure 24). XhoI was selected for use in this digestion since recognition sites for this enzyme are located on either side of the insert borders and are not present

Figure 23 PCR Generation of a Pufferfish ERM Specific Genomic Probe

PCR amplification of a 450 bp pufferfish ERM genomic fragment was carried out using gene specific primers that were designed to intron sequence lying on either side of exon 4 (AB. 12422 & AB. 12423). The template used in each of the PCR reactions was pERM-A. Once more a Mg^{2+} titration (0.5 mM - 2.5 mM) was performed. A no template negative control has also been included. A 100 bp DNA ladder has been included, and a number of size markers have been indicated. The position of the major PCR product (450 bp) has been noted.

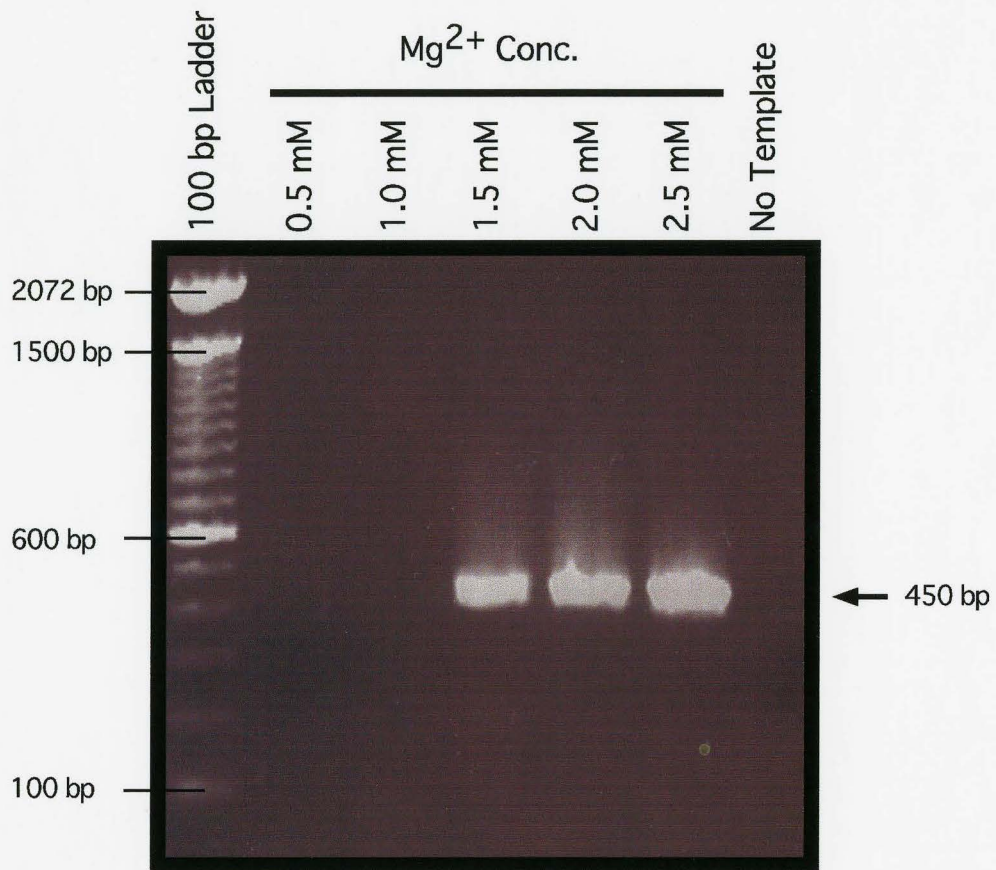
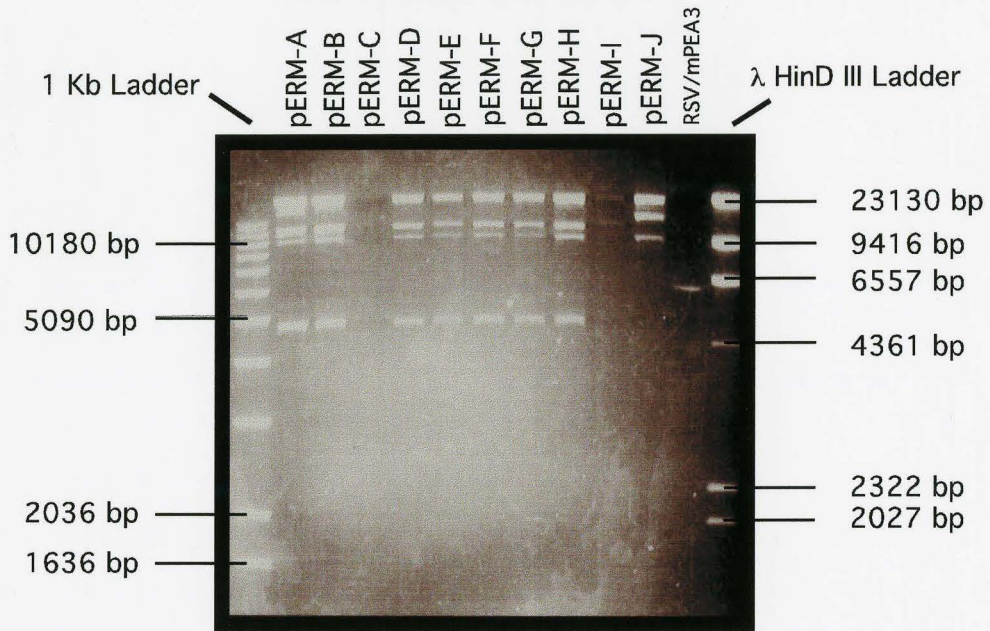


Figure 24 Restriction Digestion of Putative Positive Pufferfish λ Bacteriophage Clones

Putative positive λ bacteriophage clones pERM-A, pERM-B, pERM-C, pERM-D, pERM-E, pERM-F, pERM-G, pERM-H, pERM-I and pERM-J were digested with XhoI and the product fragments were separated by agarose gel electrophoresis. As a positive control for future southern hybridization 25 ng of the RSV/PEA3 expression vector was included during the electrophoresis. A 1 kb and a λ HindIII ladder have been included, and a number of size markers have been indicated.

XhoI



anywhere else in the EMBL SP6/T7 λ phage arms. Following restriction digestion of the ten putative positive λ clones, the product restriction fragments were separated in a 0.8% agarose gel and visualized by ethidium bromide staining. The observed restriction fragment banding pattern revealed the presence of a 9.2 and a 20.1 kb band in each of the samples except for pERM-C and pERM-I, whose sample lanes appeared not to contain DNA. The 9.2 and 20.1 kb bands correspond to the known size of the EMBL SP6/T7 λ phage arms. All of the remaining phage clones, except pERM-J, possessed the same XhoI restriction digestion banding pattern. Aside from the phage arm bands, a single restriction fragment was observed at an approximate size of 4.9 kb relative to the 1 kb ladder. A further restriction fragment migrated at a size of approximately 10 kb. Clone pERM-J exhibited a unique restriction digestion pattern, with two phage arms evident, but only a single other band running at a position corresponding to a DNA fragment of an approximate size of 15 to 16 kb.

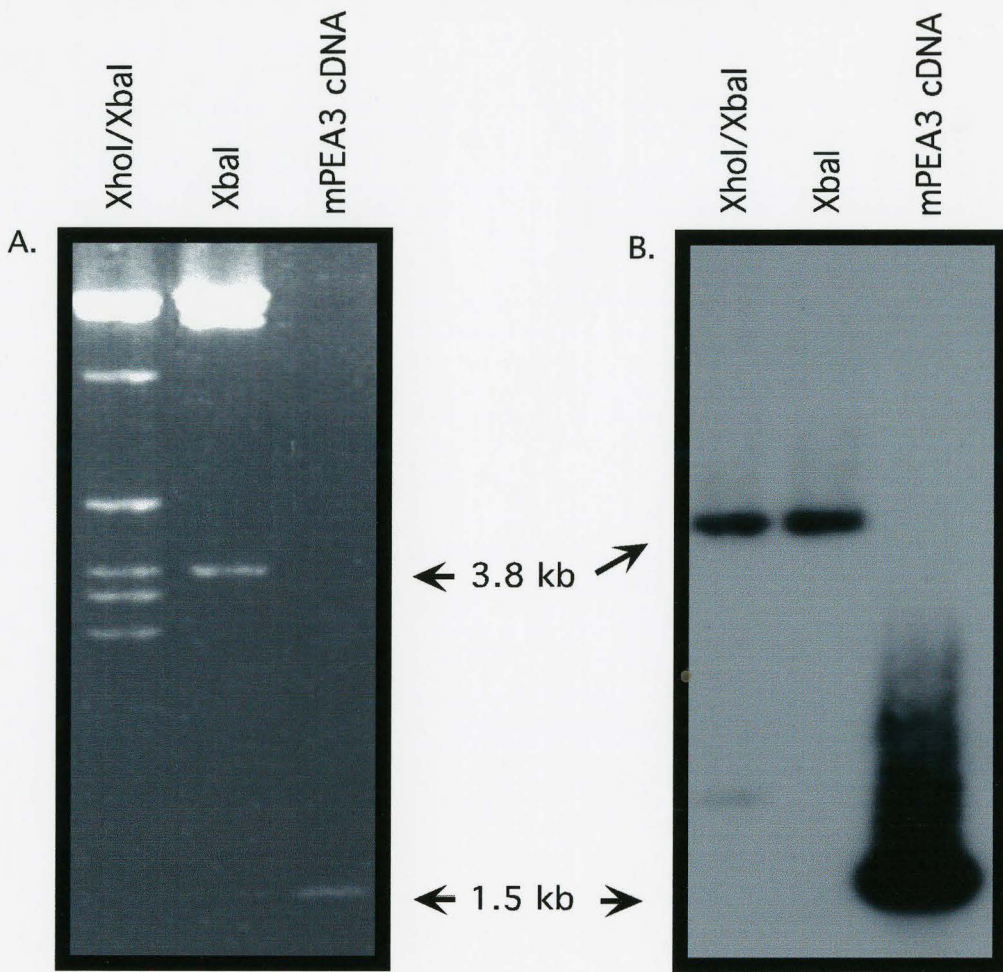
Clone pERM-A was also digested with XhoI and XbaI to identify other restriction fragments that could possibly hybridize to the full-length mouse PEA3 cDNA during southern blot analysis (Figure 25A). To this end, pERM-A was first cut in a double digestion with XhoI and XbaI followed by a single digestion with XbaI. Digestion of pERM-A with XhoI and XbaI produced the expected 20.1 and 9.2 kb λ phage arm fragments, as well as four additional digestion products of sizes 4.8, 3.8, 3.5 and 3.1 kb. An additional band was also observed to migrate at a position similar to the 20.1 kb phage arm, and presumably this band represents undigested λ phage DNA of an extremely high molecular weight. Digestion of pERM-A with XbaI alone produced three bands, one corresponding to an approximate size of 3.8 kb; and two other bands that were extremely large (greater than 20 kb). These two larger bands represent the λ phage arms fused to portions of the insert DNA. As before, another band was observed that was larger than the λ phage arm fragments, and again, this band most probably represents uncut λ phage DNA.

Restriction endonuclease digestion was also performed on the λ phage clones that were isolated from the second library screen using the genomic pufferfish ERM probe (data not shown). Both pERM-2A and pERM-2B, when digested with XhoI, produced the two expected EMBL SP6/T7 λ phage arm fragments as well as a single fragment of an approximate size of 15 to 16 kb. When pERM-2C was

**Figure 25 Restriction Digestion and Southern Blot Analysis of the Putative
Pufferfish ERM λ Clone pERM-A**

- A. The putative positive pufferfish ERM λ clone pERM-A was subjected to two restriction digestions. The first involved digestion with both XhoI and XbaI and the second involved digestion with XbaI alone. The products of these digestions were then separated by agarose gel electrophoresis. As a positive control for subsequent southern hybridization, 25 ng of full-length mouse PEA3 cDNA was electrophoresed. The position of this band has been indicated (1.5 kb). The position of a 3.8 kb fragment in each of the digestion lanes has also been shown.
- B. A full-length mouse PEA3 cDNA probe was hybridized to the immobilized restriction fragments from digestion of pERM-A with XhoI/XbaI and XbaI. The size of the restriction fragments (3.8 kb and 1.5 kb) that hybridized to the probe are indicated.

pERM-A



digested with XhoI, the 9.2 and 20.1 kb λ phage arm fragments were observed. In addition, an 8.0 kb fragment and a 7.2 kb fragment were also observed. Finally, digestion of pERM-2D with XhoI produced the expected λ phage arm fragments as well as two other restriction fragments of sizes 7.0 and 12 kb.

Southern Blot Analysis of λ Bacteriophage DNA

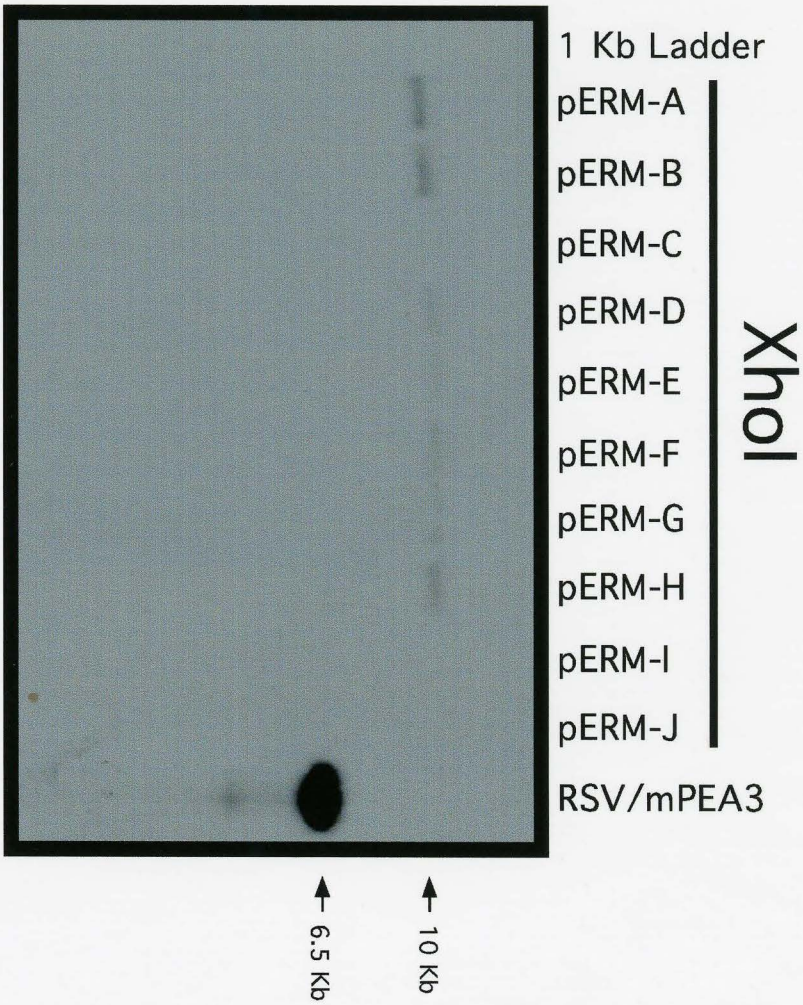
Following restriction digestion of the putative positive λ phage clones from both the first and second library screens with XhoI, transfer to nylon membranes immobilized the product restriction fragments. The membrane containing the restriction fragments from the digestion of pERM-A through pERM-J was then used for southern analysis using full-length mouse PEA3 cDNA as a probe (Figure 26). This southern analysis revealed that the probe hybridized to a single restriction digestion fragment of a size slightly larger than 10 kb. This hybridization was observed for pERM-A, pERM-B, pERM-D, pERM-E, pERM-F, pERM-G and pERM-H. Hybridization was not observed for either pERM-C or pERM-I, and it appeared that no DNA was present in these lanes. Hybridization was not observed anywhere in the pERM-J lane, and when taken together with its restriction digestion pattern, this result suggested that pERM-J was a false positive.

Southern blot analysis was also carried out on the products of the digestion of pERM-A with XhoI and XbaI, and with Xba I alone. By hybridizing a full-length mouse PEA3 cDNA to the immobilized restriction fragments from these digestions it was found that the mouse PEA3 probe specifically hybridized to the 3.8 kb band that was present in the products of each digestion (Figure 25B).

Hybridization of the genomic pufferfish ERM probe to the immobilized XhoI digestion fragments from pERM-2A and pERM-2B showed probe hybridization to a 15 -16 kb band in both samples. Hybridization was not observed to any of the bands that were produced by digestion of pERM-2C with XhoI. This result suggested pERM-2C was a false positive product. Strong probe hybridization was also observed to the 7.0 kb XhoI restriction fragment from pERM-2D (Figure 27).

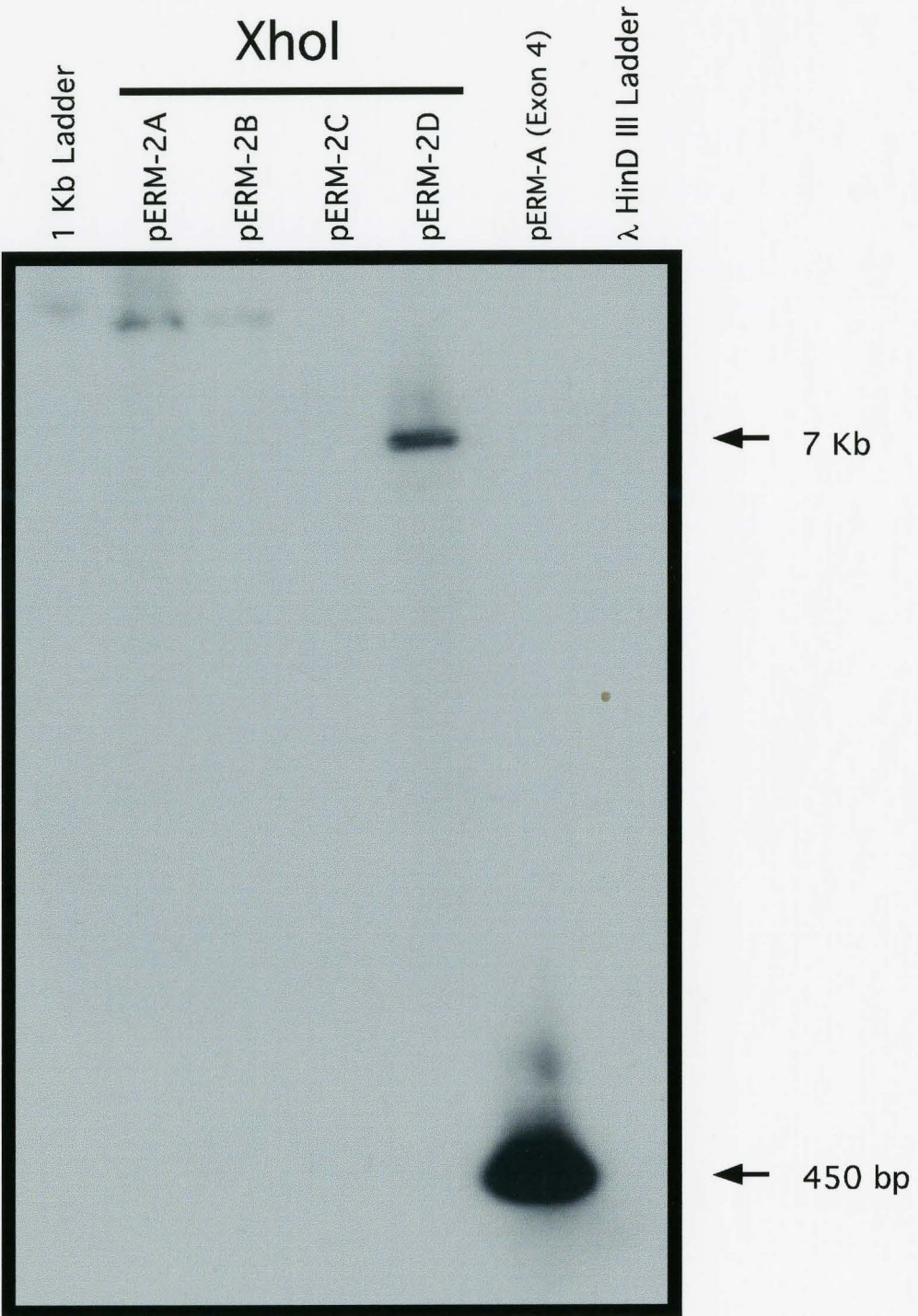
Figure 26 Southern Blot Analysis of Genomic Pufferfish λ Bacteriophage DNA from the Initial Library Screen

The full-length mouse PEA3 cDNA probe was hybridized to the immobilized XhoI restriction fragments of pERM-A, pERM-B, pERM-C, pERM-D, pERM-E, pERM-F, pERM-G, pERM-H, pERM-I and pERM-H. The size of the restriction fragment (10 kb) that hybridized to the probe in each sample lane is indicated. Hybridization is also seen to the 6.5 kb RSV/PEA3 expression vector band.



**Figure 27 Southern Blot Analysis of Genomic Pufferfish λ Bacteriophage DNA
from the Second Library Screen**

The 450 bp pufferfish ERM genomic probe containing exon 4 was hybridized to the immobilized XhoI restriction fragments of pERM-2A, pERM-2B, pERM-2C, pERM-2D. The size of the restriction fragment (7 kb) that hybridized in the pERM-2D lane is indicated. The lanes containing the immobilized 1 kb and λ HindIII DNA ladder fragments are noted.



Subcloning of DNA Fragments Containing Pufferfish ERM

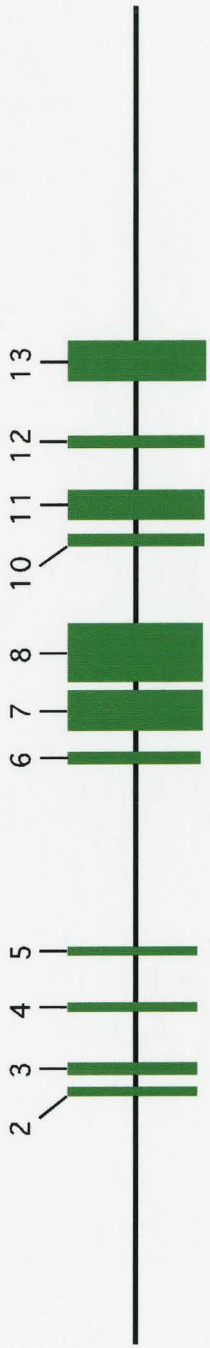
Once it had been determined that the 3.8 kb XbaI restriction fragment of pERM-A hybridized to the full-length mouse PEA3 cDNA probe during southern analysis, a subcloning strategy was undertaken. This 3.8 kb XbaI restriction fragment was purified by gel electrophoresis, followed by Qiagen Gel Extraction, and then subcloned into the pBluescript KS II cloning vector. This subcloned fragment was then used to provide sequence information from which the identity of the gene was determined. Further subcloning was also carried out by ligation of the 7.0 kb XhoI fragment (produced by digestion of pERM-2D with XhoI) with the pBluescript KS II cloning vector, which had previously been digested with XhoI. A summary of the subcloned fragments of genomic pufferfish ERM is presented in Figure 28.

DNA Sequencing of Pufferfish ERM

Subcloning of the 3.8 kb XbaI restriction digest fragment from pERM-A allowed for the production of genomic sequence spanning from 5' of exon 10 to exon 13 as well as a large portion of sequence 3' of the pufferfish ERM gene. This sequencing began by using the T3 and T7 priming sites located in the pBluescript KS II vector. Additional sequencing through the remainder of the 3.8 kb fragment was accomplished by the design of specific primers within newly defined genomic sequence, followed by the use of these primers in sequencing reactions. The entire 3.8 kb fragment was sequenced on both strands to ensure that the data were reliable, and all of the resulting sequencing information was joined together using the Sequence Editing and Analysis subprogram of the DNA Star Lasergene software package. As a result of the extremely high degree of nucleotide and amino acid identity within the Ets domain (encoded within exons 11, 12, and 13), sequencing of the 3.8 kb fragment was not, on its own, sufficient to confirm the identity of the gene. Although BLAST (Basic local alignment search tool) (Altschul et al., 1990) analysis of the genomic sequence between exons 10 and 13 confirmed that this gene

Figure 28 A Summary of Subcloned Restriction Fragments Containing Pufferfish ERM

The intron-exon organization of the pufferfish ERM gene is indicated to scale, with exons boxed in green and introns depicted by thin black lines. Each green box has been numbered according to which exon it represents. A 1 kb marker has been included to indicate scale. Below the depiction of the pufferfish ERM gene, the restriction fragments that were subcloned are indicated. On either end of the fragments the restriction enzyme recognition sites are shown.



was a PEA3 group member orthologue, the sequence was too similar to that of PEA3, ER81 and ERM to make a definitive identification. In order to identify which pufferfish PEA3 group member had been discovered, further 5' sequencing was performed to generate additional sequence outside of the Ets domain encoding exons.

Upstream genomic sequence was produced by the design of specific primers 5' of exon 10 that were used in succession in order to sequence into the intron between exons 3 and 4. This primer walking strategy used pERM-A as a template during the sequencing reactions. Once exon 4 had been found, a genomic probe was produced by PCR amplification, and this probe was used in a second library screen. Sequencing of the 7.0 kb XhoI fragment from pERM-2D (a product of the second screen) by using T3, T7 and a number of other gene specific primers revealed that the entire genomic locus resided within this subcloned fragment. The 7.0 kb XhoI subclone was then used exclusively to sequence the remainder of the gene. Using the phage clone pERM-2D, an additional 500 bp region of promoter located 5' of the XhoI 7.0 kb fragment was also sequenced using gene specific primers designed within the extreme 5' end of the 7.0 kb subcloned insert. The complete sequence of genomic pufferfish ERM is presented in Appendix F.

Computer Assisted Identification of Pufferfish ERM

Completion of the genomic sequence of this PEA3 group member also allowed for identification of the gene. The amino acid sequences corresponding to open reading frames, which bore similarity to amino acid sequences within the PEA3 group of proteins, were assembled and subjected to alignment with the mouse PEA3, ER81 and ERM amino acid sequences. By using the Multiple Sequence Alignment subprogram of the DNA Star Lasergene software package a determination of percent similarity and divergence could be determined. It was found that this pufferfish Ets gene was most closely related to mouse ERM at 44.7%, followed by mouse ER81 at 38.2%, and finally by mouse PEA3 at 33.4%. Although the percent similarities are lower than would be expected, it is important to note that this is most

likely due to the incorporation of extra amino acids into the pufferfish ERM primary amino acid sequence, necessitated by the lack of an available cDNA sequence. Despite this, the result was sufficient to identify the gene in question as being a pufferfish orthologue of the PEA3 group transcription factor ERM.

Two-way Alignment of the Human and Pufferfish ERM Promoters

Sequence Identity Between the Promoters

From the sequencing of the genomic pufferfish orthologue of the PEA3 group member ERM, approximately 1.2 kb of proximal promoter sequence was generated upstream of the putative start codon located within exon 2. Previously, a small non-coding region of the 5' end of genomic human ERM was sequenced (Monte et al., 1996). The availability of this sequence allowed for a further alignment of promoters from evolutionarily divergent Ets genes. However, the published sequence of genomic human ERM has a number of sites, contained within introns, where the sequencing is incomplete. This was particularly evident in the 5' end of the genomic sequence where approximately 2.5 kb of sequence is missing between the non-coding exon 1' and the coding exon 2.

To complete an alignment of the promoters, the human ERM sequence was split into two separate regions, one portion lying upstream of the 2.5 kb break in the sequence, and one portion lying downstream of the 2.5 kb break. Alignment of the genomic human ERM sequence upstream of this discontinuity with the pufferfish ERM sequences lying upstream of the putative start codon in exon 2 revealed that sporadic sequence identity does exist between the 5' regions of these two genes (Figure 29). However the percent similarity across the entire region under study was only approximately 20%. Further, alignment of the human ERM sequences downstream of the break to exon 4 with the pufferfish ERM sequences 3' of the putative start codon to exon 4 revealed a similar level of sequence identity (Figure 30). Once more, only

Figure 29 **Two-way Alignment of the Human and Pufferfish ERM Promoters 5' of the Sequence Break**

The human ERM genomic sequence was split into two separate regions, the first lying upstream of the 2.5 kb break in the sequence between exons 1' and 2, and the second lying downstream of the 2.5 kb break in the sequence. The human ERM sequence upstream of the discontinuity was aligned with the pufferfish ERM sequence upstream of the putative start codon in exon 2. This alignment was accomplished using the Clustal Alignment subprogram of the DNASTar software package. Two-way positional identity has been indicated in blue throughout the alignment. The defined exons within the human ERM sequence have been boxed and the exon number is indicated. No untranslated exons have been identified to date within pufferfish ERM. Putative consensus transcription factor binding sites that are found to be conserved in both promoters, and which are 100% identical to a core transcription factor consensus binding site and 85% identical or greater to the adjacent sequences within the consensus site have been indicated. The names of the transcription factors have also been included.

Pufferfish ERM Promoter **CAG**GT**TT**GTG**CC**AGTGTAAATGTGTA**GGTT**AGCATGGACG**GGCAGC**ATA
Human ERM Promoter **CAG**CC**T**-----**GGTT**-----**GGCAGC**---

Pufferfish ERM Promoter C C C A **CGGC** A **CAAA** C A G **GT** A **CA** T T T **GC** G A T T **GT** A T T T T T A **GCGT** **GG** A A G **GA**
Human ERM Promoter - - T G **CGGC** **CA** **GA** - - - **GT** **CA** G C C **GC** - - T **GT** - - - - - **GCGC** **GG** A G C **GG**
Exon 1

Pufferfish ERM Promoter A A A T T T G T T A A T T A T A T T T A C A T A C **TTCA** T **CA** C **C** A C G A A C T G T G G A G A A T
Human ERM Promoter - **TTCA** C **C** G T **C** - - - - - - - - - - - - - - - T

Pufferfish ERM Promoter **TGGGTGC** A A A T **GTTG** T **GC** T A **AA** T **CT** A T T T C A A A A G A G G G A C T G G T T A T A
Human ERM Promoter **T** **CGGA** **GC** G - - - **GTT** **CG** - **GC** C C A **GC** C - - **TTTC** -

Pufferfish ERM Promoter A T T **CA** A C T **GT** **GC** T A A A A T G **GT** **GC** T T A T A A T A A G A A A G A G **GTT** **TGA** A A T
Human ERM Promoter G C C **CA** G - - **GC** **GC** C C A **AG** G C C C **GC** T **GC** G C G C G - - - - - - - - - - - T **GCG** **TGA** G - C

Pufferfish ERM Promoter **GT** A **CT** A A T G A T A T T G A **TAC** A T A **G** A T T A A T T G A T A G A A T A T A A G G A A A **CCA**
Human ERM Promoter **G** C **GC** C - - - - - - - - - - - **TG** C **GC** C C **G** - **CCA**

Pufferfish ERM Promoter **GG** A A T A A C A A A A T T **CT** A T A G T T **GG** A T **AGA** A **CA** A A T T A A G G C A T A A T A **CAG**
Human ERM Promoter **GG** G C C G - - - - - - - - - - - **CT** G C A A - - **GG** G G **AGGA** G A G - - - - - - - - - - - - - - - **CGG**

Pufferfish ERM Promoter **CT** **G** A C A G T A T T A T **CT** **C** **CAGGTGG** C C T A T C T A C C A G T A A C A T A A A A T A C T G
Human ERM Promoter **CCG** - - - - - - - - - - - **CCT** **CAGGTGA** ↙ MyoD -

Pufferfish ERM Promoter C T A T G T G T A T G T T G T G C A T A A A A C T **GT** **GC** T **CTTT** **GTT** **G** T A T T A T T G C A C T
Human ERM Promoter - **GG** **GG** **CTTT** **GC** **G** -

Pufferfish ERM Promoter T T T T G T C T T T A A T G C T T T T G A C T G T C A A C A A C A A A **CAAG** **G** A A C A T T C T T T G
Human ERM Promoter - G G G **CAAG** **G** G G G G - - - - - - - - - -

Pufferfish ERM Promoter T A T A C **AA** **ATC** A **G** T C A C A C A C A G T **AA** **AG** T T T A **TCA** **CA** **G** T T G **C** A T A **GT** T T C A
Human ERM Promoter - - - - - **AG** **ATC** **G** - - - - - - - - - - - - - - - **AA** **AG** G G C G **TC** - **CG** **G** C C A **C** - T **G** **GT** C - - -

Pufferfish ERM Promoter **CGA** T T T T **AG** C T T **CA** T G **CA** A **CTAGA** A **GC** C C A A A T A A **GC** C A A A A A A A G G A
Human ERM Promoter **CGA** - - - - C **AG** G C C **CA** G C **C** G G **CTAGA** - **G** T T **C** - - **AC** G G **GC** T G - - - - - - - - - -

Pufferfish ERM Promoter A T C A T A **G C C** T G G G A A T T **C C C A G A** A A T C A T **G G T A** A T A **G A T T A A** A C A T **A C A G**
 Human ERM Promoter **G C T - - - G C C - - - - - C C C A G A** G C A C - C **G G G A** G G G **G G T A A A T G A C A C A G**

Pufferfish ERM Promoter A T **A** T T T T T G T A G G **G T G A A G A C T C** C T A C **G A T A T G G A T G** — Putative Start Codon
 Human ERM Promoter G C **A** - - - - - A **G T G C A G C C T G C** - - G **G G G C G G G G G**

Figure 30 Two-way Alignment of the Human and Pufferfish ERM Promoters 3' of the Sequence Break

The human ERM sequence downstream of the discontinuity was aligned with the pufferfish ERM sequence downstream of the putative start codon in exon 2 up to the 3' end of exon 4. This alignment was accomplished using the Clustal Alignment subprogram of the DNASTar software package. Two-way positional identity has been indicated in blue throughout the alignment. The defined exons with the human ERM sequence have been boxed and the exon number is indicated. Putative consensus transcription factor binding sites that are found to be conserved in both promoters, and which are 100% identical to a core transcription factor consensus binding site and 85% identical or greater to the adjacent sequences within the consensus site have been indicated. The names of the transcription factors have also been included. The putative start codon in the pufferfish ERM and the recognized start codon of human ERM or both boxed.

Pufferfish ERM Promoter	<u>A T G G A T</u> G G C T <u>T C T A T G A C C A G C A G G T T C A T T T</u> G T <u>A G C T T C G G A G A G T G T</u>
Human ERM Promoter	- <u>T A G T T</u> A A A A <u>T C T</u> - - - - <u>C C</u> - - <u>C A A A T T C A T A T T</u> A C <u>A G</u> - - - - <u>G A G G A T C C C</u>
	Putative Start Codon Exon 2
Pufferfish ERM Promoter	G A G <u>T T T C A A A G A T T T C A A G C A C A T A T G C A A T C T</u> T G G <u>A A A A G A G C T G A T T G</u>
Human ERM Promoter	<u>T T T T C C C C C A G A A A T T A C T C A A T G C T G A A A C C T C T C A A A G T G G T</u> - - <u>A T T A</u>
Pufferfish ERM Promoter	T <u>G T G A C C A G T T T T C G C T C T T C A A A C T G T A G A A A T G G C C C T C A G A A G T T G C</u>
Human ERM Promoter	- <u>G A G A C</u> - G C <u>T G A A A G C A C C A T G G A C G G G T T T T A T G A T C A G C</u> - - <u>A A G T C C C</u>
	Start Codon
Pufferfish ERM Promoter	<u>T G A G A A A T G T C C A G C A A G G C G G A A G A A G A A G T T C A T G G A C A C A G A G T T G G</u>
Human ERM Promoter	<u>T T T T A T G - G T C C C A G G G G</u> T A A <u>G T T T A T G T G G C T T T T G G T T T</u> - - <u>G T T T T G T</u>
Pufferfish ERM Promoter	<u>C C C A G G A C A C G G A A G G T A G C T G C A C C A T C T A T G - A T G A C A T G A A A C A A T G</u>
Human ERM Promoter	<u>C C T C C C T C T C C - A A T A T G A G T C T T C C C C C T G T G G A C C T C T T A C T A C A C T T</u>
Pufferfish ERM Promoter	A T A <u>C A A T G C A A T A G A T G C C A G T T A A T A A C T G A A G T A C C A A T G T G G T T A T A</u>
Human ERM Promoter	G A G <u>C C T T - C A C T T T C T G T T G G C C T C T T T C A G A A A T C T C G A T C T G A G G A A T</u>
	Exon 3
Pufferfish ERM Promoter	<u>G C A T G T G A A A C C A C C C A G A T G G A A A A A G A C A C T G G G A C A A T T A T T T T T T</u>
Human ERM Promoter	<u>G C A G A G G G C G G C C T G T G A T T G A C A G A A A G A - - - - G G A - - - - A G T T T T T</u>
Pufferfish ERM Promoter	T <u>G G C T G T A A T T G A A T A C A A C C T G A T T T C A A A A T G A A T T A G G A T A A T G A T G</u>
Human ERM Promoter	<u>G G A C A C A G A T C T G G C T C A - - - C G A T T C T G A A G</u> - <u>G T A G T A A A G C T T T C C T G</u>
Pufferfish ERM Promoter	G A C G <u>T C A C T G T G A T A G A T T C A C - C G A A A T T A A G G C C A A G A A T G G T A T T T C</u>
Human ERM Promoter	A T T A <u>T - G T T G T G G C T T C C C T G C T C C C A G T G A C A G T A G C G T G T A G A T T C T T</u>
Pufferfish ERM Promoter	T G A <u>T G A C G T G T G T T T T A A A A A T G A C T C T T G T A C A G C T T A A A T C T T A A T G T</u>
Human ERM Promoter	C C C <u>T G T C T T C T C C C T A G C G A A A G A A - - - - - A T A T C C T C A - T T C T G G G T</u>
Pufferfish ERM Promoter	<u>C G T T T C A T C A C T G T T T T G C A G A G C T C T T T C A A G A T T T A A G T C A G C T T C A A</u>
Human ERM Promoter	C - - <u>T T C T T T T C A A T T - - C A G A G C T A T T T C A G G A T C T C A G T C A A C T T C A A</u>
	Exon 4 AP-1
Pufferfish ERM Promoter	<u>G A A A T C T G G A T A G C A G A A</u>
Human ERM Promoter	<u>G A G G C T T G G T T A G C T G A A</u>

sporadic identity was observed and the percent similarity between these two sequences in this region was only 25%.

Identification of Conserved Consensus Transcription Factor Binding Sites within the Promoters

Analysis of each of the sequence alignments between human and pufferfish ERM revealed that very few transcription factor binding sites may be found to be conserved between these two sequences. It is important to indicate that a large number of consensus transcription factor binding sites were identified within the individual sequences, but that these sites simply did not overlap within the two promoters. Within the alignment which illustrates the sequence identity between the region of the human ERM sequence 5' of the 2.5 kb sequence break and the pufferfish ERM sequence 5' of the putative start codon, a single MyoD consensus binding site was found to be conserved. This site directly corresponds to the 3' border of the untranslated exon 1' in the human ERM sequence. Of further interest within this alignment was the presence of a 100% match to the consensus binding site that has been reported for PEA3 (Bowman and Hassell, private communication). However, no Ets binding sites were observed to occur at this position in the pufferfish ERM promoter following alignment. The positions of the putative MyoD and PEA3 binding sites are indicated in Figure 29.

Inspection of the alignment of the sequences 3' of the sequence break in human ERM and the sequences 3' of the putative start codon in pufferfish ERM, also indicates an absence of conserved consensus transcription factor binding sites in this region. Only one potentially conserved AP-1 binding site was found (Figure 30). This site was observed within the coding sequence of exon 4 in both the human and pufferfish ERM sequences.

Another observation that may be made regarding the pufferfish ERM promoter is that a large number of SRY binding sites may be observed within its sequence. Within the region bounded by the extreme 5' limit of the known promoter and exon 4, there are a total of 22 candidate SRY binding sites. A single SRY binding site was also found to occur in the human ERM genomic sequence 5' of exon 3.

However, none of these binding sites, within either of the ERM promoter sequences, were found to coincide following alignment, although this does not rule out the possible significance of these sites to the regulation of expression of ERM.

DISCUSSION

Isolation and Sequencing of the Chicken PEA3 Locus

Size of Genomic Chicken PEA3

Sequencing of chicken PEA3 revealed that this gene encompasses approximately 12.5 kb of the chicken genome. This is quite comparable to the size of mouse PEA3, which resides within approximately 15 kb of genomic sequence (Smillie, 1993). It is also comparable to human PEA3, which is roughly 15 kb in size (Barrett, 1997), although direct size comparison is difficult because the human gene has not been fully sequenced through a single large intron between exons 4 and 5. Within the genomic sequence, the positions of the coding exons have been determined for the 3' end of the gene. These exons were identified through sequence comparison of the 3' partial cDNA with the genomic sequence. The exact positions of exons 1 to 7 remain undetermined, and only the open reading frames bearing amino acid homology to other PEA3 orthologues in these regions have been identified. In addition to the coding sequence that has been identified, a portion of the proximal promoter region has also been sequenced. This region spans 2 kb upstream from the putative start codon of chicken PEA3.

Intron-Exon Organization of Chicken PEA3

Direct comparison of the molecular organization of introns and exons between human, mouse and chicken PEA3 was also performed. The intron-exon organization of chicken PEA3 was found to be nearly identical to that of the PEA3 orthologues from human and mouse, with one notable exception. Namely, inspection of the 3' partial cDNA of chicken PEA3, and the genomic sequence in the 3' end of the gene

revealed that exon 9 was not present. The absence of this exon in each of these sequences refutes the hypothesis that this exon was differentially spliced out of the PEA3 message. It is important to note however, that a number of open reading frames do exist between exons 8 and 10 in the genomic sequence. The possibility therefore exists that exon 9 is present in the genomic sequence, but its nucleotide sequence is so dissimilar that it cannot be identified. This, in conjunction with a hypothetical splice variant of the PEA3 message, could explain why it appears that this exon is not present. However, if the absence of exon 9 is valid, the potential importance of the amino acids encoded by it is brought into question. Two opposing viewpoints may be raised with regards to this question. Firstly, the absence of the exon may suggest that these amino acids play a trivial role in the overall folding, and therefore function, of the native protein. Secondly, it may be argued that over the course of evolution, this exon has become incorporated into the gene and has subsequently introduced functional domains into PEA3. The second hypothesis is unlikely due to the fact that in zebrafish PEA3 (Brown et al., 1998) exon 9 is present. The presence of exon 9 in an evolutionarily removed vertebrate as well as in higher vertebrates, suggests that exon 9 has most probably been lost during the evolution of the chicken genome. Therefore, exon 9 does not play a role in PEA3 function within chickens, but it may be required for the specific role PEA3 plays within mice and humans. This proposal is supported by the finding that amino acids encoded within exon 9 may function, in whole or in part, as a DNA binding inhibitory domain of mouse PEA3 (B. Bojovic and J.A. Hassell, private communication).

Conservation of Putative Phosphorylation Sites in Chicken PEA3

The ability of Ets transcription factors to bind with, and transactivate, target gene promoters is thought to be influenced by post-translational modification, specifically phosphorylation. A number of Ets factors have been shown to be targets of the mitogen activated protein kinase (MAPK) pathway. Specifically, the Ets factors Elk-1, Sap-1, and Erp/Net/Sap-2 are all activated by extracellular response

kinases (ERK) (Janknecht et al., 1993; Marais et al., 1993; Whitmarsh et al., 1995). These factors have also been shown to be phosphorylated by the Jun N-terminal kinase (JNK) as well (Cavigelli et al., 1995; Whitmarsh et al., 1995). Further, the PEA3 group Ets transcription factors ERM and ER81 have both been demonstrated to be targets of phosphorylation by intracellular signaling cascades (Janknecht et al., 1996a; Janknecht et al., 1996b). Recent work has also demonstrated that PEA3 is phosphorylated by the extracellular response kinases (ERK-1 and ERK-2) as well as by JNK (M. Cox and S. Perron, private communication). Within the primary amino acid sequence of mouse PEA3 there are eight candidate proline directed serine phosphorylation sites. Alignment of the amino acid sequence of mouse and human PEA3 demonstrated that all eight were present in these two PEA3 proteins. However, alignment of the chicken and zebrafish PEA3 protein sequences with that of the mouse and human, revealed that a subset of the candidate phosphorylation sites were not conserved. Specifically, the phosphorylation sites at serines 22, 198, 285 and 458 (numbered relative to the mouse amino acid sequence) are not found in the amino acid sequence of either chicken PEA3 or zebrafish PEA3. Of these sites only serine 458 has been found to be phosphorylated *in vivo*. This site has also been found to be phosphorylated by JNK *in vitro*. The remaining candidate sites at positions 90, 101, 134 and 143 were conserved within all four PEA3 proteins. Interestingly, sites 90, 101 and 134 were all found to undergo phosphorylation *in vivo*. These three serine residues have also been demonstrated to be phosphorylation targets of both ERK and JNK kinases. The presence of these three sites within the primary amino acid sequence of chicken and zebrafish PEA3 supports the claim that these serine residues are *bonafide* sites of phosphorylation in PEA3.

Isolation and Sequencing of the Pufferfish PEA3 Locus

Partial Size and Intron-Exon Organization of Genomic Pufferfish PEA3

Although the sequencing of genomic pufferfish PEA3 remains incomplete, a number of coding exons have been identified. Specifically, exons 2, 3, 4, 5, 12 and 13 have been sequenced and, from this sequencing, a number of observations may be made. Firstly, whereas the genomic distance between the start codon in exon 2 and the 3' end of exon 4 is 654 bp in human PEA3, and 764 bp in mouse PEA3, the comparable sequence in pufferfish PEA3 was found to be 743 bp in length. Although this finding suggests that the spatial relationships between exons in the 5' end of the PEA3 gene have been conserved, closer inspection of the sizes of the individual introns within this sequence has revealed that this is not the case. The intron between exons 2 and 3 is 191 bp in mouse PEA3 and 147 bp in human PEA3. Surprisingly, the same intron in pufferfish PEA3 is found to be 431 bp long: roughly two-fold larger than in the mouse gene, and almost three-fold larger than in the human gene. This indicates that the general genomic architecture of the 5' end of the PEA3 gene in pufferfish is not evolutionarily conserved. It may further suggest that these introns, which are not reduced in size within the pufferfish genome, may potentially contain regions of functional significance to the regulation of expression of PEA3.

Secondly, within both mouse and human PEA3 there is an extremely large intron situated between exons 4 and 5. Within genomic mouse PEA3 this intron is roughly 6 kb in length. Given the hypothesis that gene size is reduced seven to eight fold in the pufferfish genome (Brenner et al., 1993), the equivalent intron in the pufferfish PEA3 gene should be approximately 750 to 850 bp in length, however this is not the case. Combination of the genomic sequence that has been identified downstream of exon 4 and upstream of exon 5, currently totals 1392 bp, with an unknown amount of sequence still residing between the two endpoints. From this observation it is clear that the size of this intron is not reduced to the extent that was expected. This finding could suggest that the large intron between exons 4 and 5 may house regulatory

elements that govern the expression of PEA3, however, alignment of existing sequence from this intron with all available PEA3 orthologues has not revealed any substantial sequence similarity (data not shown).

Thirdly, a genomic distance comparison between PEA3 orthologues was conducted at the extreme 3' end of the gene. The distance between exon 12 and the putative polyadenylation site (AATAAA) is 1290 bp in mouse PEA3, and 1370 bp in human PEA3. Within pufferfish PEA3 there is an approximate two-fold reduction in the size of the equivalent sequence. Again, as with the intron sequences between exons 4 and 5, it appears that the 3' end of the gene does not support that there is a seven to eight fold reduction in the size of the pufferfish genome. Although the regions that have been sequenced are relatively small, it appears that pufferfish PEA3 is not reduced in size by the predicted amount, and that it may in fact be fairly large.

One potential problem with the sequencing of pufferfish PEA3 is that the genomic sequence was generated through the use of three separate cosmids, namely pPEA3-1, 019G13 and pPEA3-1A. This causes a problem, because the continuous, non-overlapping sequences must be identified as originating from the same gene. The first two cosmids, pPEA3-1 and 019G13, were used to produce the continuous sequence surrounding exon 5, and the sequence encompassing exons 12, 13, and the neighboring pufferfish homologue of human RNA helicase 1 (HRH1). The 900 bp XmaI fragment from 019G13 was found to overlap with the intron sequence directly 5' of exon 5 from cosmid pPEA3-1. Thus, both of these sequences can be identified as pufferfish PEA3 genomic sequence, based on both sequence similarity to other PEA3 orthologues, and synteny conservation with the human PEA3 locus. The problem arises when an attempt is made to properly identify the genomic sequence of the 5' end of the gene. The continuous sequence surrounding exons 2, 3 and 4, including the promoter sequence, was produced by subcloning a 6.5 kb BamHI fragment from the cosmid pPEA3-1A. At this point none of the genomic sequence from this cosmid overlaps with the sequence that was produced from pPEA3-1 or 019G13. This raises the possibility that the genomic sequence inclusive of exons 2 to 4 is not PEA3 but rather a PEA3 homologue, most probably ER81 (since pufferfish ERM has already been sequenced). However, the 5' RACE product sequence was found to be 100% identical to the exon 5 sequence found in the continuous genomic

sequence produced from cosmids pPEA3-1 and 019G13. Further, the cDNA sequence of exons 2, 3 and 4 was also found to be 100% identical to the corresponding exons found within the genomic sequence produced from pPEA3-1A. Based on these findings, it is highly unlikely that the 5' genomic sequence inclusive of the promoter is from a pufferfish version of ERM or ER81, although the remainder of the genomic sequence of pufferfish PEA3 must be completed to further verify this claim. To accomplish this, further subcloning and sequencing from pPEA3-1A and from 019G13 should be performed in order to sequence the large intron between exons 4 and 5. Also, further attempts should be made at producing cDNA sequence between exons 5 and 13 using techniques such as 3' RACE or RT-PCR. This type of approach would allow for direct sequencing of pPEA3-1 using specific primers designed within exons 6, 7, 8, 9, 10 and 11.

Conservation of Chromosomal Synteny at the PEA3 Locus

Another characteristic of the pufferfish genome that has become evident is the fact that genomic synteny, or the order relationships between genes, is often conserved between higher vertebrates and pufferfish (Brenner et al., 1993; Trower et al., 1996). In humans, approximately 1 Mb of the BRCA1 locus on chromosome 17q21 has been sequenced (Brody et al., 1995; Osborne-Lawrence et al., 1995), and it was found that human PEA3 (E1AF) resides directly 5' to the human RNA helicase HRH1 gene, a homologue of the yeast RNA helicase Prp22 (Ono et al., 1994). Sequencing of the 5' end of pufferfish PEA3 revealed that a pufferfish homologue of HRH1 is situated 100 bp directly 3' of the putative pufferfish PEA3 polyadenylation signal. The discovery of the HRH1 homologue 3' of the pufferfish PEA3 gene provides further evidence to support the claim that this pufferfish Ets gene is in fact PEA3. From this sequencing it is also possible to determine that pufferfish PEA3 and the pufferfish homologue of HRH1 lie in opposite directions in the pufferfish genome. This is analogous to the situation in the human genome where these two genes are found to be transcribed in opposite directions (Brody et al., 1995). Thus, sequencing of the

pufferfish PEA3 genomic locus has provided yet another example of conservation of chromosomal synteny between the pufferfish and higher vertebrate genomes. This finding lends further support to the use of the pufferfish genome as a model for gene discovery within extremely large genomes.

Identification of a 5' Partial cDNA of Pufferfish PEA3

In addition to the genomic sequencing that has been accomplished for pufferfish PEA3, a partial 5' cDNA has also been produced by 5' RACE. This product was generated through the use of pufferfish PEA3 specific primers derived from exon 5 sequence and poly (A)⁺ heart RNA. The isolation of a PEA3 cDNA using heart RNA is surprising since PEA3 mRNA expression is highly restricted, and has not been identified in the mouse heart (Xin et al., 1992). The expression of PEA3 in the adult pufferfish heart suggests that the role of PEA3 in the pufferfish may be significantly different from that which it plays in the adult mouse, however, further study of the overall expression pattern of PEA3 in the adult pufferfish is required to validate this claim.

Alignment of the deduced amino acid sequence (from the 5' partial cDNA) of pufferfish PEA3 with other PEA3 orthologues (Figure 31), as well as with the equivalent sequences from mouse ERM, mouse ER81, human ERM, human ER81 and xenopus ER81 has shown that the amino-terminal end of pufferfish PEA3 is 92.5% homologous to the amino-terminus of zebrafish PEA3. Further, pufferfish PEA3 was found to be 61.3% homologous with the amino-terminal ends of both mouse and human PEA3. When the amino-terminal end of pufferfish PEA3 was compared to that of mouse ERM and mouse ER81 the percent similarities were found to be 51.4% and 53.3% respectively. Comparable percent similarities were observed in comparisons with human ERM and human ER81, where the similarities were found to be 50.0% and 53.3% respectively. The amino terminal end of the pufferfish PEA3 protein was also compared to the equivalent sequence from the newly defined *Xenopus laevis* ER81 protein (Chen et al., 1999; Munchberg et al., 1999). This comparison revealed a percent similarity of 56.0% (Table 3). When the

**Figure 31 Alignment of the Deduced Partial Amino Acid Sequence of Pufferfish PEA3
with Other PEA3 Group Proteins**

Alignment of the deduced amino acid sequence from the pufferfish PEA3 5' RACE product with the amino-terminal ends of various evolutionarily removed PEA3 group proteins. The corresponding exon borders have been indicated within the sequence. The amino acids contained within the conserved acidic domain are underlined in red. Positional identity in all seven sequences is indicated in blue.

Table 3 **Percent Similarity and Percent Divergence Between the Amino Terminal End of Pufferfish PEA3 and Other PEA3 Group Proteins**

This table summarizes the percent similarity and percent divergence between the amino-terminus of pufferfish PEA3 and the amino-termini of other PEA3 group proteins from a range of evolutionarily removed species.

Percent Similarity

		1	2	3	4	5	6	7	8	9		
Percent Divergence	1		90.2	61.3	61.3	44.6	43.2	45.3	45.3	46.7	1	Mouse PEA3
	2	9.8		61.3	61.3	45.9	44.6	45.3	45.3	48.1	2	Human PEA3
	3	32.5	32.5		92.5	51.4	50.1	53.3	53.3	57.3	3	Zebrafish PEA3
	4	32.5	32.5	7.5		51.4	50.1	53.3	53.3	56.1	4	Pufferfish PEA3
	5	45.9	44.6	41.1	41.1		94.6	67.6	68.9	64.9	5	Mouse ERM
	6	47.3	45.9	42.5	42.5	5.4		68.9	70.3	67.6	6	Human ERM
	7	49.3	49.3	43.2	43.2	28.4	27.1		97.3	80.1	7	Mouse ER81
	8	49.3	49.3	43.2	43.2	27.1	25.7	2.7		78.7	8	Human ER81
	9	48.1	46.7	39.2	40.5	31.1	28.4	20.0	21.3		9	Xenopus ER81
	1	2	3	4	5	6	7	8	9			

existing cDNA sequence of pufferfish PEA3 was compared to the genomic sequence of pufferfish ERM, the nucleotide sequences were found to be dissimilar. This indicates that this newly identified putative PEA3 gene was not the pufferfish version of ERM. These results, when taken together, suggest that the cDNA produced during the 5' RACE analysis was in fact from the pufferfish orthologue of PEA3 and not one of the other PEA3 group members ERM or ER81.

Alignment of PEA3 Orthologue Promoters

A large number of genomic PEA3 orthologue genes are beginning to emerge from a wide variety of evolutionarily removed vertebrate species. These genes can be extremely useful for understanding which cis-acting elements within non-coding sequence govern the expression of PEA3. As was discussed earlier, the utility of direct promoter comparison has been substantiated in a number of systems, most notably *Hoxb-4* (Aparicio et al., 1995), *Otx2* (Kimura et al., 1997), *Wnt-1* (Rowitch et al., 1998) and *CRABP-1* (Kleinjan et al., 1998). In each of these cases, comparison of the promoter from the Japanese Pufferfish *Fugu rubripes* gene with the promoter from the mouse gene has revealed the importance of specific elements to the expression of their respective genes. Following these examples a four-way alignment of the PEA3 promoters from mouse, human, chicken and pufferfish was conducted.

Observed Sequence Identity

Alignment of the mouse and human PEA3 promoters showed an extremely large amount of sequence identity spanning from -300 bp upstream of the transcriptional start site to the intron between exons 1' and 2. This region was found to be 90% identical between the human and mouse PEA3 genes (Barrett, 1997). The occurrence of sequence conservation within introns downstream of the transcriptional

start site is not unprecedented, as this situation has been reported within a number of other genes, including the distantly related Ets gene PU.1 (Chen et al., 1995). The inclusion of additional promoter sequences was required to further narrow the search for regulatory elements within the PEA3 promoter.

From the four-way promoter alignment it is evident that the observed sequence similarity exists in a much smaller region than was previously described between the mouse and human genes (Figure 22). This region aligns directly upstream of the start site of transcription relative to the mouse promoter. Specifically, this region was found to extend from position +1 to approximately -260 bp upstream of the transcriptional start site. Downstream of the start site of transcription there does not appear to be any further remarkable sequence identity, however, directly 3' of exon 3 there does appear to be a short stretch of sequence that is conserved between all of the genomic PEA3 orthologues under study. When the remainder of the promoter alignment is examined for the occurrence of four-way nucleotide identity it is apparent that only sporadic conservation exists outside of coding exons 3 and 4 and the +1 to -260 region described above. This finding suggests that the region bounded by position +1 and -260 relative to the mouse PEA3 promoter plays a role in the regulation of PEA3 expression. A detailed analysis of the consensus transcription factor binding sites found within this region may serve to indicate those transcription factors, which may govern this expression.

Conserved Transcription Factor Binding Sites

Analysis of the promoters using the Transcription Factor Database shows that there are a number of different consensus transcription factor binding sites present within the region spanning from +1 to -260. Further inspection of this small region has also revealed that a number of consensus binding sites are only conserved across three of the four promoters. These sites include: 2 HNF3 β , 2 NFY, 1 SRY, 1 TCF,

1 AP-2, and 1 v-myb. Outside of the +1 to -260 region a single AP-2 binding site was found to be located within exon 1 in the mouse, human and chicken genomic sequences. A single c-Ets-1 binding site was also found directly downstream of exon 1 in the mouse, human, and chicken sequences.

A limited number of consensus transcription factor binding sites were also observed in each of the four promoters. Specifically, centered at position -248 in the context of the mouse promoter, a 15 bp region of sequence identity was observed with only a single position not conserved across the four promoters under study. When this 15 bp region of the pufferfish PEA3 promoter is analyzed using the transcription factor database, a consensus SRY binding site is revealed. However, analysis of the equivalent 15 bp region of the mouse, human and chicken PEA3 promoters did not reveal similar findings. In fact, no binding sites were found within these sequences at all. Nonetheless, it is likely that this 15 bp element within the mouse, human and chicken promoters does serve as a binding site for the SRY transcription factor, regardless of the fact that a point mutation may have occurred during the evolution of these promoters. It is also unlikely that this 15 bp element has simply been spared from random mutational events. A more plausible explanation for the conservation of this element is that this sequence serves some definite purpose.

A second binding site that was conserved across all four promoters was a single AP-1 transcription factor binding site located approximately 100 bp upstream of the transcriptional start site in the mouse PEA3 promoter. This site was perfectly aligned in the mouse, human and chicken promoters, however the pufferfish AP-1 binding site was slightly offset in the 5' direction. The final binding site that was found to be conserved across all four promoters, but was located outside of the +1 to -260 region, overlapped with the extreme 3' border of exon 3. This sequence was found to correspond to a consensus binding site for the transcription factor δ EF1. It is important to note that the binding sites that have been described above are only putative and that they should not be taken as *bonafide* transcription factor binding sites until proven experimentally. Also, there may be additional sites within these promoters that were not identified, possibly because the binding site is not defined by the transcription factor database.

At the extreme 5' boundary of the +1 to -260 region, a single HNF3 β binding site was found to be conserved between the mouse, human and pufferfish PEA3 promoters. A second HNF3 β site was also found to occur approximately 100 bp upstream of the transcriptional start site in the mouse, human and chicken PEA3 promoters. The forkhead winged helix transcription factor, the founding member of the HNF3/forkhead family, has been shown to control gut and central nervous system development in *Drosophila* (Beck et al., 1999). HNF3/forkhead homologues have been identified in both mice and humans and have been shown to be expressed in a wide variety of tissues including: brain, kidney, liver, intestine, lung, lymphocytes, and testis (Monaghan et al., 1993; Wiese et al., 1997; Frank & Zoll, 1998; Beck et al., 1999). This pattern of expression is of interest since PEA3 expression has been observed at varying levels within a number of these tissues, specifically within the brain, kidney, intestine and testis (M. Laing, private communication). Also, the expression of a human transcription factor (Human HNF3/forkhead-like 5 - HFKL5), homologous to members of the HNF3/forkhead factor family, has been observed in fully differentiated neurons in fetal and adult brain, as well as in parasympathetic ganglia of the small intestine (Wiese et al., 1997). Recent evidence has shown that the PEA3 group of transcription factors, most notably ER81 and PEA3, are expressed in motor and sensory neurons (Lin et al., 1998). PEA3 and ERM have also been proposed to act downstream of neuroregulin-1, a member of the EGF superfamily, which influence growth and differentiation of neuronal cells (Hagedorn et al., in press). Of further interest is the fact that HNF3 factors (α , β and γ) have been shown to bind to the neuronal promoter of the human aromatic L-amino decarboxylase gene, thus indicating that HNF3 factors are expressed in neuronal cells (Raynal et al., 1998). The potential role that HNF3/forkhead factors play in transactivating the PEA3 promoter in neuronal cells is interesting since it has recently been hypothesized that PEA3 is required during innervation of the male sex organs (M. Laing, private communication). In fact, PEA3-null mice are phenotypically sterile (Laing, 1998), possibly due to this aberrant innervation. The observed co-expression of members of the HNF3/forkhead family and PEA3 in a variety of tissues may suggest that the reported binding sites for HNF3 β in the PEA3 promoter may be functionally significant.

Two well-conserved putative binding sites have also been identified for the Y chromosome associated HMG (High Mobility Group) box containing SRY transcription factor. These sites are located within the +1 to -260 region. The first, being present at the extreme 5' end, is found in all four promoter sequences and the second, being found approximately 100 bp upstream of the transcriptional start site, is conserved in the mouse, human and chicken PEA3 promoters. SRY acts as a molecular switch that initiates the pathway that leads to differentiation of the genital ridge to testis rather than ovaries in the developing male embryo (Brennan et al., 1998). Interestingly, PEA3 has also been shown to be expressed in the genital ridge of developing embryos (M. Laing, private communication). Further, SRY is expressed in both fetal and adult Sertoli and germ cells in the testis (Hacker et al., 1995; Salas-Cortes et al., 1999). PEA3 has been shown to be expressed at moderate to high levels within the testis. It is therefore possible that SRY is capable of upregulating PEA3 expression in the testis via the conserved binding sites within the promoter. A further major site of expression of PEA3 mRNA is in the adult epididymis. The SRY related SOX9 transcription factor has been shown to be expressed at high levels in both the adult testis and the epididymis (Kent et al., 1996). The fact that SOX9 is expressed in the adult epididymis raises the possibility that this factor may upregulate expression of PEA3 through one of the two SRY consensus binding sites.

Interestingly, the SRY related SOX transcription factors may also play a role in regulating PEA3 expression within the mammary gland as well. During embryogenesis PEA3 expression is lost in the mammary gland of male mice (L. MacNeil, private communication). The SRY transcription factor has been shown to act as either a transcriptional activator or a transcriptional repressor depending upon the target promoter. Specifically, SRY has been shown to repress the P450 aromatase promoter and to transactivate the Mullerian inhibiting substance (MIS) promoter. P450 aromatase catalyzes the conversion of testosterone to estradiol during female gonadal specification, while MIS, a member of the transforming growth factor β family, is responsible for regression of female reproductive ducts in male mice. Therefore, by functioning in a dual role, SRY is capable of acting as the molecular switch leading to development of male specific structures (Haqq et al., 1993). The SRY related factor SOX4 has been shown to be expressed in both normal and breast cancer cells (Graham et al., 1999). The presence of SRY binding sites within the

PEA3 promoter could suggest that SRY or SOX factors could possibly act in a repressive role within the mammary gland of male mice. Promoter mapping studies in FM3A cells using deletions of the mouse PEA3 promoter have provided evidence to support this claim. Specifically, the region of the mouse PEA3 promoter between -256 and -156 has been shown to decrease expression from luciferase reporter constructs (Barrett, 1997). A conserved SRY binding site may be found at position -250 within the mouse PEA3 promoter (Figure 22). This provides evidence to suggest that SRY binding sites in the PEA3 promoter may potentially act in a repressive role.

Two NFY (CAAT binding factor) binding sites were found within the mouse, human and chicken PEA3 promoters, with the first being present at position -191, and the second being found at position -135 (numbered relative to the mouse promoter). NFY is a heterotrimeric transcription factor that has been shown to interact with the central region of the co-activator p300. The formation of this complex has been shown to be greatly increased by the presence of cAMP. It has been proposed that the interaction of NFY with p300 provides a mechanism by which promoters lacking cAMP-response-elements (CRE) respond to cAMP. Further, NFY has been shown to interact with TFIIB, which is a component of the basal transcription machinery (Faniello et al., 1999).

A single AP-1 consensus binding site was found to be conserved between all four of the PEA3 promoters under study. This binding site was found at position -97, and was perfectly aligned in the human, mouse and chicken PEA3 promoters. A single AP-1 binding site was also found in the pufferfish PEA3 promoter, although this site was situated slightly 5' of the others in the four-way alignment. The occurrence of a conserved AP-1 binding site in the PEA3 promoter is significant because AP-1 binding sites have been found to occur in close proximity to PEA3 binding sites in other promoters. One such promoter is the urokinase-type plasminogen activator (uPA) promoter in which two PEA3/AP-1 sites cooperate to mediate transcriptional activation following induction by TPA and FGF-2 (D'Orazio et al., 1997). The presence of a conserved AP-1 binding site in the proximal PEA3 promoter is of interest because footprinting analysis of the mouse PEA3 promoter has indicated that PEA3 binds to a position (-49 to -27) just 3' of the conserved

AP-1 binding site (Benz et al., 1997). This finding may suggest that PEA3 and AP-1 cooperate to transactivate the PEA3 promoter.

A single conserved TCF (T cell factor)/LEF-1 (Lymphoid enhancer factor-1) binding site was found to be conserved within the mouse, human and chicken PEA3 promoters. TCF/LEF-1 transcription factors mediate the nuclear response to Wnt signals by heterodimerization with β -catenin (Eastman & Grosschedl, 1999). Once this interaction has occurred, the β -catenin/TCF complex may bind with and transactivate target promoters. Specific target genes that are known to be responsive to β -catenin mediated signaling include cyclin D1 (Tetsu & McCormick, 1999; Shtutman et al., 1999) and matrylisin (Crawford et al., 1999). In fact, the expression of each of these genes is known to be upregulated in colon carcinoma cell lines where there is an increase in free stabilized β -catenin levels. Multiple intestinal neoplasia (MIN) mice bearing mutations in both alleles of the adenomatous polyposis coli (APC) gene (Mahmoud et al., 1999) have been shown to overexpress PEA3 in colorectal tumors (H. Crawford, private communication). This suggests that the presence of a conserved TCF binding site within the PEA3 promoter may confer responsiveness to both the Wnt signaling pathway and to uncontrolled accumulation of β -catenin as a result of mutational events leading to cellular transformation.

Two AP-2 binding sites were also found to be conserved between the mouse, human and chicken PEA3 promoters. However, throughout the region of sequence conservation (from +1 to -260) no AP-2 binding sites were found in the pufferfish PEA3 promoter. The two conserved AP-2 sites were found to lie on either side of the major transcriptional start site located at the beginning of exon 1. AP-2 binding sites are of interest to the expression of PEA3 since AP-2 (α , β and γ) transcription factor levels have been found to be increased in cell lines overexpressing c-erbB-2 (*Her2/neu*) and c-erbB-3 (Kraus et al., 1987; Skinner & Hurst, 1993). High levels of PEA3 mRNA have been observed in *Her2/neu* induced mammary tumors (Trimble et al., 1993). These findings suggest that increased levels of AP-2 may play a role in the overexpression of PEA3. Other sites of co-expression of the AP-2 transcription factors and erbB receptor family members exist. Specifically, AP-2 factors and the erbB2 and erbB3 receptor tyrosine kinases are expressed in the mouse cerebellum (Ozaki et al., 1998). Since PEA3 expression has been observed at a

high level in the brain, it is possible that AP-2 sites within the PEA3 promoter may govern PEA3 expression following activation of signaling cascades downstream of erbB2. Further, erbB2 and erbB3 expression has also been reported to be enhanced in colorectal tumors (Maurer et al., 1998). Since PEA3 expression is upregulated in MIN mouse colorectal tumors, it may be possible that this overexpression is a consequence of both β -catenin/TCF activity as well as other transcription factors, possibly including AP-2, which are responsive to erbB2 signaling. PEA3 expression has also been reported in the mouse uterus and intestine (M. Laing, private communication). The presence of erbB receptor family members in each of these tissues (Lim et al., 1998; Kataoka et al., 1998) may further suggest that PEA3 transcriptional regulation may be controlled by AP-2 transcription factors here as well.

A single consensus v-myb binding site also appeared to be conserved in the PEA3 promoter, and was located directly 3' to the start site of transcription in exon 1 in the mouse, human and chicken sequences. A single v-myb binding site was also present in the pufferfish PEA3 sequence, however, it was located slightly 5' of the v-myb site that was conserved in the other three sequences. Ets transcription factors and c-myb have been found to have opposing actions on the promoters of human and mouse colony-stimulating factor-1 receptor (c-fms) genes. In particular, a number of Ets factors, including c-Ets-1, c-Ets-2, PU.1 and PEA3 were found to transactivate the c-fms proximal promoter, while the c-myb proto-oncogene was found to repress promoter activity in macrophage and block the transactivating function of c-Ets-1 and c-Ets-2 (Reddy et al., 1994). These findings suggest that, the v-myb binding site in the PEA3 promoter may serve a repressive role in the context of PEA3 expression.

Of particular interest to the study of the PEA3 promoter is the search for potentially conserved Ets binding sites. In the past it has been shown that PEA3 is capable of transactivating its own promoter in COS cells. As was previously mentioned, PEA3 is also capable of binding to the mouse promoter, thereby inducing DNase-I hypersensitivity sites between positions -49 and -27, and positions -302 and -280 in the mouse PEA3 promoter (Benz et al., 1997). These findings provide convincing evidence in support of the claim that PEA3 autoregulates its own expression. Identification of conserved Ets/PEA3 binding sites within the four PEA3 promoters would be extremely beneficial to the understanding of this phenomenon,

specifically because it might allow for the identification of the PEA3 responsive element. Interestingly, the PEA3 footprint that was found between positions -49 and -27 falls within the region of homology defined by the four-way promoter alignment. Although this footprint site did not contain a consensus Ets binding site, the possibility still exists that this region of the PEA3 promoter may be PEA3 responsive. A survey of the promoter alignment revealed only a single site where c-Ets-1 binding sites (and therefore a potential PEA3 binding site) were conserved across a number of the PEA3 promoters. In particular a c-Ets-1 binding site was found directly 3' of exon 1 in the mouse, human and chicken PEA3 promoter sequences. A c-Ets-1 binding site was also found in the pufferfish PEA3 promoter at a position slightly upstream of the sites identified within the other promoter sequences. Whereas the c-Ets-1 binding site found in the mouse PEA3 promoter possessed the prototypical GGAA/T core binding motif, the c-Ets-1 binding sites found in the other three promoter sequences possessed only a GGA core element. The conservation of this site may also indicate that other members of the Ets transcription factor family, aside from PEA3, are capable of binding to, and transactivating, the PEA3 promoter. Other potential Ets binding sites were also found throughout each of the individual promoters, however they did not occur within the regions of sequence identity. It is possible then that other Ets binding sites function as binding sites for PEA3, but that the locations of these sites vary within the promoters.

A single candidate binding site for the DNA binding protein δ EF1 was also found to reside directly 3' of exon 3 in each of the PEA3 promoters being studied. δ EF1 contains a homeodomain and two zinc finger clusters, and is believed to be the vertebrate homologue of the *Drosophila* factor *zfh-1* (zinc finger homeodomain-containing factor-1) (Takagi et al., 1998). This factor has previously been shown to regulate eye lens specific transcription by binding to the δ 1-crystallin enhancer (Funahashi et al., 1993). This transcription factor is also present in a number of other mouse tissues including restricted expression in a few sites within the brain and spinal cord (Takagi et al., 1998). In addition to its role as a transcriptional activator, δ EF1 has also been shown to carry out a repressive function by competitively binding to E2 box elements (Sekido et al., 1996). The E2 box (5'-CACCTG-3') or related sequences, have been shown to act as binding sites for members of the myogenic basic helix-loop-helix (bHLH)

transcription factor family (Sekido et al., 1994). Interestingly, PEA3 has been shown to be an important regulator of activated satellite cell function following muscle degeneration (Taylor et al., 1997). The up-regulation of PEA3 in activated myoblasts is co-incident with the accumulation of MyoD in the same cells and it is therefore possible that myogenic factors affect the expression of PEA3 during myblast activation. However, PEA3 expression is down-regulated in fully differentiated myotubes (Taylor et al., 1997) and this might indicate a possible role of the transcriptional repressor δ EF1. It is possible that δ EF1 may act competitively with myogenic factors to modulate PEA3 expression during different stages of myogenic stem cell activation and differentiation.

Common Features within the PEA3 Promoters

Alignment of the proximal PEA3 promoters from human, mouse, chicken and pufferfish has revealed a number of interesting features that may be involved in governing both transcription of the various PEA3 genes, as well as the translation of PEA3 mRNA in each system. Firstly, the start site of transcription has been determined for both the mouse (Smillie, 1993) and human (King et al., database submission) PEA3 (E1AF) sequence and in each case the major start site is centered on a well-conserved initiator element or cap signal (Bucher, 1990). Inspection of the chicken PEA3 sequence following alignment, indicated that the chicken sequence also possessed a close match to the initiator element consensus. This finding suggests that the thymidine residue in the chicken initiator might function as the major start site of transcription for the chicken PEA3 gene. The start site of transcription in the pufferfish PEA3 gene has not been identified, nor does a candidate start site align with the other PEA3 sequences. As a result, it is not possible to infer the position of a potential transcriptional start site in pufferfish. Further 5' RACE experiments are required to produce the extreme 5' end of the pufferfish PEA3 cDNA. In the mouse gene, a second transcriptional start site was found to reside 3' of the major start point (exon 1') (Smillie,

1993), however this start site does not appear to be conserved across all four promoters. Given that a second transcript originating at a second transcriptional start site has not been identified in human PEA3, these results suggest that the first start site of initiation located at the beginning of exon 1 is responsible for transcriptional initiation of the majority of PEA3 transcripts.

Residing slightly upstream (position -22) of exon 1 and of the major transcriptional start site is a putative CCAAT box, which shows a high degree of conservation across the mouse, human and chicken promoters. The CCAAT box sequence within these promoters is highly similar to the previously reported CCAAT box consensus sequence (Bucher, 1990). In vertebrates, this promoter element is found between positions -212 and -57 relative to the transcriptional start site. Since the majority of CCAAT boxes lie within this specified range, it is possible that the putative PEA3 CCAAT box lies too close to the start site of transcription. However, even though the majority of CCAAT boxes are found to lie slightly 5' of position -57, there have been a small number reported outside of this optimal range (Bucher, 1990). Transcription factors which are known to bind to the CCAAT element include NF-Y and CP-1. As previously discussed, NFY is a heterotrimeric transcription factor that has been shown to interact with both the co-activator p300, and TFIIB, a component of the basal transcription machinery. CP-1 has been shown to bind with high affinity to the CCAAT elements contained within the human α -globin promoter and the adenovirus major late promoter (Chodosh et al., 1988).

Further inspection of the four-way promoter alignment revealed sequence identity surrounding the start site of translation. Specifically, the mouse, human and pufferfish PEA3 genes were each found to possess two putative start codons within exon 2. Chicken PEA3, was found to contain only one putative ATG start codon. Alignment of the promoters, including exons within the 5' end of the PEA3 genes, indicates that the chicken PEA3 start codon aligned with the second start codon of human and mouse PEA3. A similar alignment was not observed between the mouse, human, chicken and pufferfish PEA3 sequence because the pufferfish start codon aligns further upstream than that of the other three genes. The second ATG start codon in all four sequences was found to possess an adenosine residue at position -3 relative to the ATG. It has been proposed that this configuration surrounding the start codon is sufficient

for start codon selection during translation (Kozak, 1991). When taken together, the absence of the first ATG in chicken PEA3, and the discovery of a conserved adenosine residue at position -3 upstream of the second start codon in human, mouse and pufferfish PEA3, suggests that the biologically relevant start codon is most likely the second of the two.

Missing Features within the Pufferfish PEA3 Promoter

In addition to providing useful information regarding sequence conservation, the four-way PEA3 promoter alignment also highlights regions that are missing within one or more of the promoter sequences. For instance, whereas both the mouse and human PEA3 genomic sequences possess a CA repeat sequence between exons 1 and 1' (+221 to +421), the chicken and pufferfish genomic sequences do not. Given that CA repeats have been shown to act as silencer elements within promoters (Wu et al., 1994), their presence in the mouse and human genomic sequence, might suggest that the CA element is important to the transcriptional regulation of PEA3 in higher vertebrates.

It is also important to note that the putative start codon in exon 2 of pufferfish PEA3 was found to align upstream from the position where the other three start codons were found. This suggests that the sequence, essentially from the start of exon 1 up to the 5' end of exon 2, in the mouse, human and chicken sequence is missing in the pufferfish. One interpretation of this finding is that this sequence plays a minimal role in the regulation of PEA3 expression. Another is that the inclusion of this sequence in higher vertebrate genomes functions to add a higher level of transcriptional regulation to these genes. Since mouse PEA3 is known to possess two transcriptional start sites (Smillie, 1993), it is reasonable to predict that there may be two unique promoters driving transcription from each site. The absence of the second promoter element from the start of exon 1 up to the beginning of exon 2 may then suggest that this region of the mouse, human, and chicken genomic PEA3 sequence is responsible for a highly restricted, tissue specific expression of PEA3 transcripts bearing exon 1' as an untranslated exon. In fact, promoter mapping

studies have shown that the non-coding sequence of the mouse PEA3 promoter between positions +1 and +676 was capable of producing maximal reporter gene activity in the FM3A mouse mammary tumor cell line (Barrett, 1997). This finding could suggest that the second promoter was active in this cell line, while the first promoter that drives expression from exon 1, was silent. It is also important to note that when the non-coding sequences from -156 to +676 were assayed for promoter activity in the FM3A cell line, a similar level of reporter gene activity was found, however, this may be due to the fact that the sequences between -156 and +1 are transcriptionally inactive in FM3A cells. Further experimentation involving other PEA3 expressing cell lines is required to assess the ability of the sequences upstream of exon 1 to support transcription. Such experiments would also shed light on the possible existence of a dual promoter.

Isolation and Sequencing of the Pufferfish ERM Locus

Size of Genomic Pufferfish ERM

Sequencing of pufferfish ERM revealed that this gene resides within approximately 4.5 kb of pufferfish genomic sequence. Although a substantial portion of the gene sequence is unknown (Monte et al., 1996) the approximate size of the human ERM gene is roughly 65 kb. This indicates that the pufferfish orthologue of the ERM gene is reduced in size by approximately 14 fold. This finding was unexpected since it has been reported that, on average, pufferfish genes are reduced in size by seven to eight fold in comparison to their counterpart genes in higher vertebrates (Brenner et al., 1993). In addition to the coding sequence that has been identified, a portion of the proximal promoter region has also been sequenced. This region spans approximately 1.2 kb upstream of the putative start codon of pufferfish ERM.

Intron-Exon Organization of Pufferfish ERM

Both pufferfish and human ERM share a similar molecular organization of introns and exons. As was observed in chicken PEA3, exon 9 is apparently not present, however this finding cannot be completely confirmed since a cDNA sequence is not available for pufferfish ERM. It is important to indicate that a number of open reading frames exist within the genomic sequence between exons 8 and 10. As a result, exon 9 may be present, but its nucleotide sequence (and deduced amino acid sequence) may be too dissimilar from other ERM orthologues for an identification to be made. The absence of exon 9 in both chicken PEA3 and pufferfish ERM possibly suggests that the amino acids encoded by this exon are relatively unimportant to the function of these two proteins. However, as was argued earlier, the possibility exists that the incorporation of exon 9 into higher vertebrate versions of PEA3 and ERM may have introduced functional domains into these proteins. Thus, the possibility remains that in chicken, and in pufferfish, the PEA3 group of transcription factors may have evolved without exon 9.

Two-way Alignment of the Human and Pufferfish ERM Promoters

Sequence Identity Between the Promoters

The availability of promoter sequence for the Ets transcription factor ERM, from two evolutionarily divergent species (human and pufferfish) made it possible to identify conserved regions within these promoters, as well as specific transcription factor binding sites that might govern the expression of ERM. This kind of information is extremely valuable to the understanding of the expression of the related Ets transcription factor PEA3, since these two genes are found to be co-expressed in a variety of tissues during mouse embryonic development (Chotteau-Lelievre et al., 1997).

Since the 5' end of the genomic human ERM sequence is discontinuous in a number of locations (particularly within large introns) two separate alignments were produced. The first illustrated the two-way sequence identity that was found upstream of a break in the human ERM sequence between exons 1' and 2. This alignment showed an extremely sporadic pattern of sequence identity with no significantly long stretches of sequence conservation between the two promoters. This alignment also showed that the two sequences were only 20% identical to one another. When the sequences lying downstream of the sequence break in human ERM were aligned with a region of pufferfish ERM sequence encompassing exons 2 to 4, a similar level of sequence similarity was observed (approximately 25%). These two alignments demonstrate that additional human ERM promoter sequence is required to make such a comparison useful. It is important to note that the sequence break between exons 1' and 2 in human ERM may house regulatory elements that are important to the transcriptional regulation of ERM in a similar manner to that reported for PU.1, where regulatory elements were found to reside downstream of the transcriptional start site (Chen et al., 1995). An analogous situation has also been reported for the mouse PEA3 promoter, where the sequences from +1 to +676 relative to the transcriptional start site was found to constitute a core promoter (Barrett, 1997). In order to more fully examine this promoter, it will be necessary to complete the 5' end of the genomic sequence of human ERM, and to determine additional promoter sequence from other organisms.

Identification of Conserved Consensus Transcription Factor Binding Sites Within the Promoters

Within the promoter alignments there was a surprising lack of consensus transcription factor binding sites conserved between the two promoters. In fact, only two specific binding sites were found to be present within both of the promoters. The first was a conserved MyoD binding site that was found to correspond to the 3' border of the untranslated exon 1'. The existence of a MyoD binding site within the ERM promoter is not surprising since there is evidence to support a role for PEA3 in the activation,

migration, and differentiation of myogenic stem cells (Taylor et al., 1997). The fact that there is no apparent phenotype involving muscle dysfunction in PEA3-null mice (M. Laing, private communication) may suggest that ERM is able to functionally substitute for PEA3 in these mice, thereby allowing a normal phenotype to emerge.

The second binding site that was found to be conserved was a consensus AP-1 binding site. It has been reported that Ets transcription factors may function in tandem with AP-1 to effect transactivation of target promoters (D'Orazio et al., 1997). A number of Ets binding sites were found in the human ERM promoter, including a PEA3 binding site that was 100% identical to the consensus PEA3 binding site. Based on this result, the possibility exists that Ets factors and AP-1 may function cooperatively to transactivate the ERM promoter.

Unfortunately, the lack of human ERM promoter sequence has hindered the search for conserved consensus transcription factor binding sites. Further sequencing of the human ERM gene, as well as the isolation of additional ERM genes would allow for a more substantial analysis of evolutionary sequence conservation within their promoters.

CONCLUSION

The genomic orthologues of PEA3 from the chicken and Japanese Pufferfish *Fugu rubripes* were isolated from genomic libraries. The chicken PEA3 gene was completely sequenced, including approximately 2 kb of its proximal promoter, while the pufferfish PEA3 gene was only partially sequenced, including approximately 1 kb of its proximal promoter. These promoter sequences were aligned with the known promoters of both the human and mouse PEA3 genes. This alignment revealed that a region between positions +1 and -260 relative to the transcriptional start site in the mouse PEA3 gene was conserved across all four promoters. Further analysis identified a number of candidate transcription factor binding sites present within this defined region. The conservation of these binding sites suggests that they may play a distinct role in governing the expression of PEA3, although further experimentation is required to validate the importance of these sites. These findings could in the future guide in the selection of candidate promoter sequences to be used to drive expression of β -galactosidase in transgenic mice. These experiments could possibly identify the minimal core PEA3 promoter and its essential transcription factor binding sites. An additional outcome of the search for pufferfish PEA3 was the discovery of a pufferfish orthologue of the related Ets transcription factor ERM. Alignment of the proximal promoter from the pufferfish ERM gene with the existing human ERM promoter sequence failed to identify a clearly conserved region within the promoter. Further sequencing of both the human and pufferfish ERM promoters, as well as the isolation of a mouse ERM promoter, will be necessary to elucidate the mechanism of transcriptional regulation of ERM in higher vertebrates.

REFERENCES

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., & Lipman, D.J. (1990). *J. Mol. Biol.*, 215: 403-410.
- Aparicio, S., Morrison, A., Gould, A., Gilthorpe, J., Chaudhuri, C., Rigby, P., Krumlauf, R., & Brenner, S. (1995). *Proc. Natl. Acad. Sci.*, 92: 1684-1688.
- Bargmann, C.I., Hung, M-C., & Weinberg, R.A. (1986). *Nature*, 319: 226-230.
- Barrett, J. (1997) M.Sc. Thesis, McMaster University, Hamilton, Ontario.
- Beck, S., Sommer, P., dos Santos, S.E., Blin, N., & Gott, P. (1999). *DNA Cell Biol.*, 18: 157-164.
- Beitel, G.J., Tuck, S., Greenwald, I., & Horvitz, H.R. (1995). *Genes & Dev.*, 9: 3149-3162.
- Ben-David, Y., Giddens, E., Letwin, K., & Bernstein, A. (1991). *Genes & Dev.*, 5: 908-918.
- Benz, C.C., O'Hagan, R.C., Richter, B., Chang, C.H., Xiong, X., Chew, K., Ljung, B.M., Edgerton, S., Thor, A., & Hassell, J.A. (1997). *Oncogene*, 15: 1513-1525.
- Birnboim, H.C. (1983). *Methods Enzymol.*, 100: 243-255.
- Bories, J.C., Willeford, D.M., Grevin, D., Davidson, L., Camus, A., Martin, P., Stehelin, D., & Alt, F.W. (1995). *Nature*, 377: 635-638.
- Braun, B.S., Frieden, R., Lessnick, S.L., May, W.A., & Denny, C.T. (1995). *Mol. Cell Biol.*, 15: 4623-4630.
- Brennan, J., Karl, J., Martineau, J., Nordqvist, L., Schmahl, J., Tilmann, C., Ung, K., & Capel, B. (1998). *J. Exp. Zool.*, 281: 494-500.
- Brenner, S., Elgar, G., Sandford, R., Macrae, A., Venkatesh, B., & Aparicio, S. (1993). *Nature*, 366: 265-268.
- Brody, L.C., Abel, K.J., Castilla, L.H., Couch, F.J., McKinley, D.R., Yin, G., Ho, P.P., Merajver, S., Chandrasekharappa, S.C., Xu, J., Cole, J.L., Struewing, J.P., Valdes, J.M., Collins, F.S., & Weber, B.L. (1995). *Genomics*, 25: 238-247.
- Brown, L.A., Amores, A., Schilling, T.F., Jowett, T., Baert, J-L., de Launoit, Y., Sharrocks, A.D. (1998). *Oncogene*, 17: 93-104.
- Brown, T.A., & McKnight, S.L. (1992). *Genes & Dev.*, 6: 2502-2512.
- Bucher, P. (1990). *J. Mol. Biol.*, 212: 563-578.
- Burt, D.W., Bumstead, N., Bitgood, J.J., Ponce de Leon, F.A., & Crittenden, L.B. (1995). *Trends Genet.*, 11(5): 190-194.

- Chang, C-H., Scott, G.K., Kuo, W-L., Xiong, X., Suzdaltseva, Y., Park, J.W., Sayre, P., Erny, K., Collins, C., Gray, J.W., & Benz, C.C. (1997). *Oncogene*, 11: 1549-1560.
- Cavigelli, M., Dolfi, F., Claret, F-X., & Karin, M. (1995). *EMBO*, 14(23): 5957-5964.
- Chen, H-M., Ray-Gallet, D., Zhang, P., Hetherington, C.J., Gonzalez, D.A., Zhang, D., Moreau-Gachelin, F., & Tenen, D.G. (1995). *Oncogene*, 11: 1549-1560.
- Chen, J.H., Vercamer, C., Li, Z., Paulin, D., Vandebunder, B., & Stehelin, D. (1996). *Oncogene*, 13(8): 1667-1675.
- Chen, Y., Hollemann, T., Grunz, H., & Pieler, T. (1999). *Mechanisms of Development*, 80: 67-76.
- Chodosh, L.A., Baldwin, A.S., Carthew, R.W., & Sharp, P.A. (1988). *Cell*, 53: 11-24.
- Chotteau-Lelievre, A., Desbiens, X., Pelczar, H., Defossez, P.A., & de Launoit, Y. (1997). *Oncogene*, 15: 937-952.
- Crawford, H.C., Fingleton, B.M., Rudolph-Owen, L.A., Heppner Goss, K.J., Rubinfeld, B., Polakis, P., & Matrisian, L.M. (1999). *Oncogene*, 18: 2883-2891.
- Dalton, S., & Treisman, R. (1992). *Cell*, 68: 597-612.
- D'Orazio, D., Besser, D., Marksitzer, R., Kunz, C., Hume, D.A., Kiefer, B., & Nagamine, Y. (1997). *Gene*, 201: 179-187.
- Delattre, O., Zucman, J., Plougastel, B., Desmaze, C., Melot, T., Peter, M., Kovar, H., Joubert, I., de Jong, P., Rouleau, G., Aurias, A., & Thomas, G. (1992). *Nature*, 359: 162-165.
- de Launoit, Y., Baert, J-L., Chotteau, A., Monte, D., Defossez, P-A., Coutte, L., Pelczar, & H., Leenders, F. (1997). *Biochemical and Molecular Medicine*, 61: 127-135.
- de Launoit, Y., Audette, M., Pelczar, H., Plaza, S. & Baert, J. (1998). *Oncogene*, 16: 2065-2073.
- Dittmer, J. & Nordheim, A. (1998). *Biochimica et Biophysica Acta*, 1377: F1-F11.
- Donaldson, L., Petersen, J.M., Graves, B., & McIntosh, L.P. (1996). *EMBO*, 15: 125-134.
- Eastman, Q., & Grosschedl, R. (1999). *Current Opinion in Cell Biology*, 11: 233-240.
- Elgar, G., Sandford, R., Aparicio, S., Macrae, A., Venkatesh, B., & Brenner, S. (1996). *Trends Genet.*, 12(4): 145-150.
- Faniello, M.C., Bevilacqua, M.A., Condorelli, G., de Crombrughe, B., Maity, S.N., Avvedimento, V.E., Cimino, F., & Costanzo, F. (1999). *J. Biol. Chem.*, 274: 7623-7626.
- Feinberg, A.P., & Vogelstein, B. (1983). *Anal. Biochem.*, 132: 6-13.
- Frank, S., & Zoll, B. (1998). *DNA and Cell Biology*, 17(8): 679-688.
- Funahashi, J., Sekido, R., Muair, K., Kamachi, Y., & Kondoh, H. (1993). *Development*, 119: 433-446.

- Goebel, M.G., Moreau-Gachelin, F., Ray, D., Tambourin, P., Tavitian, A., Klemsz, M.J., McKercher, S.R., Celeda, A., Van Bervern, D., & Maki, A. (1990). *Cell*, 61: 1165-1166.
- Golay, J., Introna, M., & Graf, T. (1988). *Cell*, 55: 1147-1158.
- Golub, T.R., Barker, G.F., Lovett, M., & Gilliland, D.G. (1994). *Cell*, 77: 307-316.
- Golub, T.R., Barker, G.F., Bohlander, S.K., Hiebert, S.W., Ward, D.C., Bray-Ward, P., Morgan, E., Raimond, S.C., Rowley, J.D., & Gilliland, D.G. (1995). *Proc. Natl. Acad. Sci.*, 92: 4917-4921.
- Graham, J.D., Hunt, S.M., Tran, C., & Clarke, C.L. (1999). *J. Mol. Endocrinol.*, 22(3): 295-304.
- Guy, C.T., Webster, M.A., Schaller, M., Parsons, T.J., Cardiff, R.D., & Muller, W.J. (1992). *Proc. Natl. Acad. Sci.*, 89: 10578-10582.
- Hacker, A., Capel, B., Goodfellow, P., & Lovell-Badge, R. (1995). *Development*. 121: 1603-1614.
- Hagedorn, L., Mercader, N., Paratore, C., Suter, U., & Sommer, L., in press.
- Haqq, C.M., King, C-H., Donahue, P.K., & Weiss, M.A. (1993). *Proc. Natl. Acad. Sci.*, 90: 1097-1101.
- Higashino, F., Yoshida, K., Fujinaga, Y., Kamio, K., & Fujinaga, K. (1993). *Nucleic Acids Research*, 21(3): 547-553.
- Higashino, F., Yoshida, K., Noumi, T., Seiki, M., & Fujinaga, K. (1995). *Oncogene*, 10: 1461-1463.
- Ichikawa, H., Shimizu, K., Hayashi, Y., & Ohki, M. (1994) *Cancer Res.*, 54: 2865-2868.
- Isobe, M., Yamagishi, F., Yoshida, K., Higashino, F., & Fujinaga, E. (1995). *Genomics*, 28: 357-359.
- Janknecht, R., & Nordheim, A. (1993). *Biochimica et Biophysica Acta*, 1155(3): 346-356.
- Janknecht, R. (1996a). *Mol. Cell. Biol.*, 16(4): 1550-1556.
- Janknecht, R., Monte, D., Baert, J-L., & de Launoit, Y. (1996b). *Oncogene*, 13: 1745-1754.
- Jeon, I-S., Davis, J.N., Braun, B.S., Sublett, J.E., Roussel, M.F., Denny, C.T., & Shapiro, D.N. (1995). *Oncogene*, 10: 1229-1234.
- Johnson, P.F., & McKnight, S.L. (1989). *Annu. Rev. Biochem.*, 58: 799-839.
- Karim, F.D., Urness, L.D., Thummel, C.S., Klemsz, M.J., McKercher, S.R., Celada, A., van Beveren, C., Maki, R.A., Gunther, C.V., Nye, J.A., & Graves, B.J. (1990). *Genes & Dev.*, 4: 1451-1453.
- Kataoka, H., Joh, T., Kasugai, K., Okayama, N., Moriyama, A., Asai, K., & Kato, T. (1998). *Life Sciences*, 63(7): 553-564.
- Kaya, M., Yoshida, K., Higashino, F. Mitaka, T., Ishii, S., & Fujinaga, K. (1996). *Oncogene*, 12: 221-227.
- Kent, J., Wheatley, S.C., Andrews, J.E., Sinclair, A.H., & Koopman, P. (1996). *Development*. 122: 2813-2822.

- Kimura, C., Takeda, N., Suzuki, M., Oshimura, M., Aizawa, S., & Matsuo, I. (1997). *Development*, 124: 3929-3941.
- Kleinjan, D.A., Dekker, S., Guy, J.A., & Grosveld, F.G. (1998). *Transgenic Research*, 7:85-94.
- Klemsz, M.J., McKercher, S.R., Celada, A., van Beveren, C., & Maki, R.A. (1990). *Cell*, 61: 113-124.
- Kozak, M., (1991). *J. Biol. Chem.*, 266: 19867-19870.
- Kraus, M.H., Popescu, N.C., Amsbaugh, S.C., & King, C.R. (1987). *EMBO*, 6: 605-610.
- Laing, M. (1998). Ph.D Thesis, McMaster University, Hamilton, Ontario.
- LaMarco, K.L., Thompson, C.C., Byers, B.P., Walton, E.M., & McKnight, S.L. (1991). *Science*, 253: 789-792.
- Laudet, V., Niel, C., Duterque-Coquilland, M., LePrince, D., Stehelin, D. (1993). *Biochemical and Biophysical Research Communications*, 190: 8-14.
- Laudet, V., Hanni, C., Stehelin, D., & Duterque-Coquilland, M. (1999). *Oncogene*, 18: 1351-1359.
- LePrince, D., Gegonne, A., Coll, J., DeTaisne, C., Schneeberger, A., Lagrou, C., & Stehelin, D. (1983). *Nature*, 306: 395-397.
- Lim, H., D, S.K., & Dey, S.K. (1998). *Developmental Biology*, 204: 97-110.
- Lin, J., Saito, T., Anderson, D.J., Lance-Jones, C., Jessell, T.M., & Arber, S. (1998). *Cell*, 95: 393-407.
- Liu, D., Pavlopoulos, E., Moid, W., Moschonas, N., & Mavrothalassitis, G. (1997). *Oncogene*, 14: 1445-1451.
- Lopez, M., Oettgen, P., Akbarali, Y., Dendorfer, U., & Libermann, T. (1994). *Mol. Cell. Biol.*, 14: 3292-3309.
- Macleod, k., LePrince, D., & Stehelin, D. (1992). *Trends in Biochemical Sciences*, 17: 251-256.
- Mahmoud, N.N., Bilinski, R.T., Churchill, M.R., Edelmann, W., Kucherlapati, R., & Bertagnolli, M.M. (1999). *Cancer Res.*, 59(2): 353-359.
- Mao, S., Frank, R.C., Zhang, J., Miyazaki, Y., & Nimer, S.D. (1999). *Mol. Cell. Biol.*, 19: 3635-3644.
- Marais, R., Wynne, J., & Treisman, R. (1993). *Cell*, 73: 381-393.
- Martin, M.E., Piette, J., Yaniv, M., Tang, W-J., & Folk, W.R. (1988). *Proc. Natl. Acad. Sci.*, 85: 5839-5843.
- Matrisian, L.M. (1994) *Ann. N.Y. Acad. Sci.*, 732: 42-50.
- Maurer, C.A., Friess, H., Kretschmann, B., Zimmermann, A., Stauffer, A., Baer, H., Korc, M., & Buchler, M. (1998). *Human Pathology*, 29(8): 771-777.

- May, W.A., Lessnick, S.L., Braun, B.S., Klemsz, M., Lewis, B.C., Lunsford, L.B., Hromas, R., & Denny, C.T. (1993). *Mol. Cell. Biol.*, 13: 7393-7398.
- Miklos, G.L. & Rubin, G.M. (1996). *Cell*, 86: 521-528.
- Mitchell, P.J., & Tjian, R. (1989). *Science*, 245(4916): 371-378.
- Monaghan, A. P., Kaestner, K.H., Grau, E., & Schutz, G. (1993). *Development*, 119: 567-578.
- Monte, D., Baert, J., Defossez, P.A., de Launoit, Y., & Stehelin, D. (1994). *Oncogene*, 9: 1397-1406.
- Monte, D., Coutte, L., Baert, J., Angeli, I., Stehelin, D., & de Launoit, Y. (1995). *Oncogene*, 11: 771-779.
- Monte, D., Coutte, L., Dewitte, F., Defossez, P.A., Coniat, M., Stehelin, D., Berger, R., & de Launoit, Y. (1996). *Genomics*, 35: 236-240.
- Muller, W.J., Sinn, E., Pattengate, P.K., Wallace, R., & Leder, P. (1988). *Cell*, 54: 105-115.
- Munchberg, S.R., & Steinbeisser, H. (1999). *Mechanisms of Development*, 80: 53-65.
- Muthusamy, N., Barton, K., & Leiden, J.M. (1995). *Nature*, 377: 639-642.
- Nunn, M.F., Seeburg, P.H., Moscovici, C., & Duesberg, P.H. (1983) *Nature*, 306: 391-395.
- Nunn, M.F., & Hunter, T. (1989). *J. Virol.*, 63: 398-402.
- Nye, J.A., Petersen, J.M., Gunther, C.V., Jonsen, M.D., & Graves, B. (1992). *Genes & Dev.*, 6: 975-990.
- Oettgen, P., Akbarali, Y., Boltax, J., Best, J., Kunsch, & Libermann, T. (1996). *Mol. Cell. Biol.*, 16: 5091-5106.
- Ono, Y., Ohno, M., & Shimura, Y. (1994). *Mol. Cell. Biol.*, 14: 7611-7620.
- Osborne-Lawrence, S., Welsch, P.L., Spillman, M., Chandrasekharappa, S.C., Gallardo, T.D., Lovett, M., & Bowcock, M. (1995). *Genomics*, 25: 248-255.
- Ozaki, M., Kishigami, S., & Yano, R. (1998). *Neurosci. Res.*, 30(4): 351-354.
- Peter, M., Couturier, J., Pacquement, H., Michlon, J., Thomas, G., Magdelenat, H., & Delattre, O. (1997). *Oncogene*, 14: 1159-1164.
- Petermann, R., Mossier, B.M., Aryee, D., Khazak, V., Golemis, E.A., & Kovar, H. (1998). *Oncogene*, 17: 603-610.
- Rao, V.N., Papas, T.S., & Reddy, E.S.P. (1987). *Science*, 237: 635-639.
- Rao, V.N., Huebner, K., Isobe, M., Ar-Rushdi, A., Croce, C.M., & Reddy, E.S.P. (1989). *Science*, 244: 66-70.
- Ray, D., Bosselut, R., Ghysdael, J., Mattei, M-G., Tavitian, A., & Moreau-Gachelin, F. (1992). *Mol Cell Biol.*, 12: 4297-4303.
- Raynal, J.F., Dugast, C., Le Van Thai, A., & Weber, M.J. (1998). *Brain Res. Mol. Brain Res.*, 56: 227-237.

- Reddy, E.S.P., Rao, V.N., & Papas, T.S. (1987). *Proc. Natl. Acad. Sci.*, 84: 6131-6135.
- Reddy, M.A., Yang, B.S., yue, X., Barnett, C.J., Ross, I.L., Sweet, M.J., Hume, D.A., & Ostrowski, M.C. (1994). *J. Exp. Med.*, 180: 2309-2319.
- Rorth, P., Nerlov, C., Blasi, F., & Johnsen, M. (1990) *Nucleic Acids Res.*, 18: 5009-5017.
- Rowitch, D.H., Echelard, Y., Danielian, P.S., Gellmer, K., Brenner, S., & McMahon, A.P. (1998). *Development*, 125: 2735-2746.
- Salas-Cortes, L., Jaubert, F., Barbaux, S., Nessmann, C., Bono, M.R., Fellous, M., McElreavey, K., & Roseblatt, M. (1999). *Int. J. Dev. Biol.*, 43: 135-140.
- Sambrook, J., Fritsch, E.F., & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Second Edition. Cold Spring Harbor Laboratory Press.
- Scott, E.W., Simon, M.C., Anastasi, J., & Singh, H. (1994). *Science*, 265:1573-1577.
- Sekido, R., Takagi, T., Okanami, M., Moride, H., Yamamura, M., Higashi, Y., & Kondoh, H. (1996). *Gene*, 173: 227-232.
- Seth, A., Ascione, R., Fisher, R.J., Mavrothalassitis, G.J., Bhat, N.J., & Papas, T.S. (1992). *Cell Growth & Differentiation*, 3: 327-334.
- Slamon, D.J., Clark, G.M., Wong, S.G., Levin, W.J., Ullrich, A., & McGuire, W.L. (1987). *Science*, 235: 177-182.
- Shtutman, M., Zhurinsky, J., Simcha, I., Albanese, C., D'Amico, M., Pestell, R., & Ben-Ze'ev, A. (1999). *Proc. Natl. Acad. Sci.*, 96: 5522-5527.
- Skinner, A., & Hurst, H.C., (1993). *Oncogene*, 8(12): 3393-3401.
- Smillie, D. (1993). M.Sc. Thesis, McMaster University, Hamilton, Ontario.
- Southern, E. (1975). *J. Mol. Biol.*, 98: 503-517.
- Taylor, J.M., Dupont-Versteegden, E.E., Davies, J.D., Hassell, J.A., Houle, J.D., Gurley, C.M., & Peterson, C. (1997). *Mol Cell Biol*. 17: 5550-5558.
- Takagi, T., Moribe, H., Kondoh, H., & Higashi, Y. (1998). *Development*, 125: 21-31.
- Tetsu, O., & McCormick, F. (1999). *Nature*, 398: 422-426.
- Thompson, C.B., Wang, C-Y., Ho, I-C., Bohjanen, P.R., Petryniak, B., June, C., Miesfeldt, S., Zhang, L., Nabel, G.J., Karpinski, B., & Leiden, J.M. (1992). *Mol. Cell. Biol.*, 12: 1043-1053.
- Trimble, M.S., Xin, J-H., Guy, C.T., Muller, W.J., & Hassell, J.A. (1993). *Oncogene*, 8: 3037-3042.
- Trower., M.K., Orton, S.M., Purvis, I.J., Sanseau, P., Riley, J., Christodoulou, C., Burt, D., Gee See, C., Elgar, G., Sherrington, R., Rogaev, E., St.George-Hyslop, P., Brenner, S., & Dykes, C.W. (1996). *Proc. Natl. Acad. Sci.*, 93: 1366-1369.

- Urano, F., Umezawa, A., Hong, W., Kikuchi, H., & Hata, J. (1996). *Biochemical and Biophysical Research Communications*, 219: 608-612.
- Wasylyk, B., Hahn, S.L., & Giovane, A. (1993). *European Journal of Biochemistry*, 211: 7-18.
- Watson, D.K., McWilliams, M.J., Lapis, P., Lautenberger, J.A., Schweinfest, C.W., & Papas, T.S. (1988). *Proc. Natl. Acad. Sci.*, 85: 7862-7866.
- Whitmarsh, A.J., Shore, P., Sharrocks, A.D., & Davis, R.J. (1995). *Science*, 269: 403-407.
- Wiese, S., Emmerich, D., Schroder, B., Murphy, D.B., Grzeschik, K.H., Van Kessel, A.G., & Thies, U. (1997). *DNA Cell Biol.*, 16: 165-171.
- Woods, D.B., Ghysdael, J., & Owen, M.J. (1992). *Nuc. Acids Res.*, 20: 699-704.
- Wu, T., Ikezono, T., Angus, C.W., & Shelhamer, J.H. (1994). *Nucleic Acids Res.*, 22: 5093-5098.
- Xin, J-H., Cowie, A., Lachance, P., & Hassell, J.A. (1992). *Genes & Dev.*, 6: 481-496.
- Zucman, J., Melot, T., Desmaze, C., Ghysdael, J., Plougastel, B., Peter, M., Zucker, J., Triche, T.J., Sheer, D., Turc-Carel, C., Ambros, P., Combaret, V., Lenoir, G., Aurias, A., Thomas, G., & Delattre, O. (1993). *EMBO*, 12: 4481-4487.

Appendix A Chicken PEA3 Genomic Sequence

The genomic sequence of chicken PEA3, including 2 kb of sequence upstream of the putative start codon in exon 2, is depicted. The genomic sequence has been translated according to the three possible reading frames, and the resultant amino acid sequence has been shown below each nucleotide sequence line. The deduced amino acid sequences that are homologous to other PEA3 orthologue protein sequences, and that correspond to open reading frames within the genomic sequence have been boxed. The boxes are colored and correspond to the legend that may be found at the bottom of the sequence. The putative start codon in exon 2, and the putative polyadenylation signal have been indicated. The putative stop codon is also shown. Where two possible exons are found to reside within one large open reading frame, the amino acids encoded by each exon are approximated based on amino acid sequence similarity to other PEA3 proteins. Exons 8 to 13 have been underlined in the nucleotide sequence based on the available 3' cDNA sequence.

AAAGGACTTGGACTCCCATGACCAGAAGAGCAAAAGCAAAAGGCTGCAGGAACAGCAGCAATCCACAGAGCTGGAATGTTTCATGAGGCTGATTTAGAATGACACAAAAACACTCAAAAGTGT 123
K G L G L P . P E E Q K Q K A A G T A A I H R A G M F M R L I . N D T K T L K S V
K D L D S H D Q K S K S K R L Q E Q Q Q S T E L E C S . G . F R M T Q K H S K V F
Q R T W T P M T R R A K A K G C R N S S N P Q S W N V H E A D L E . H K N T Q K C
TCCTGGGTGAGGATCTGCATTCTTTTCAGATATTCATCAACTTAAGGATGAAGATGGCTAATATCGACAGCTGAAGAGTAGAGTCAGCACCTTATTCAACACCTCCCAACCCAGGGAACAACG 246
S W V R I C I L S D I P S T . G . R W L I S T A E E . S Q H L I H N L P T Q G T T
P G S G S A F F Q I F H Q L K D E D G . Y R Q L K S R V S T L F T T S P P R E Q R
F L G Q D L H S F R Y S I N L R M K M A N I D S . R V E S A P Y S Q P P H P G N N
GGGCATCAGAGAACGGCTGAACCTCAGTCCTGATGGTACAAAACCTGAGCTCTTACCCAGCTGTTTATTTTATAAAAACCTGAGGTCGCTTATGCTCTATTCTAAACGTGTCTCTGCACCTC 369
G H Q R T A E L S P D G T K P E L F T Q L F I F I K L Q V C L C S I L N V S L H L
G I R E R L N S V L M V Q N L S S S P S C L F L . N C R S A Y A L F . T C L C T S
G A S E N G . T Q S . W Y K T . A L H P A V Y F Y K T A G L L M L Y S K R V S A P
TGCACTTCAGTCAAGAAAAGAGTCAAGAAAGCTGGGAAGGTTATTTTTGTGTACAGAGATGGGGAAGTGAAGACAACCATGCAGAGACATAAGGCCTCTGTCAAGTTCTGTCAATGCCACT 492
C T S V K K R V K K A G K V I F V Y R D G E V R R Q P C R D I R P L S V L S L P T
A L Q S R K E S R K L G R L F L C T E M G K . E D N H A E T . G L C Q F C H C P L
L H F S Q E K S Q E S W E G Y F C V Q R W G S E K T T M Q R H K A S V S S V I A H
GGATGGGAAATTAGAACATCATAAAGCAAAATAAAGTGAAGTCCCATCCGGTTTCAGAGCGCAGCCATATGCTGAAATCAGATCTGAAATAACAAAGGTTCCATTATATCTTAACC 615
G W E I R T S . S K Q I . Q W K I P S G S E R S H M L K S D L K . Q R F H Y I L T
D G K L E H H K A N K Y N S G K S H P V Q S A A I C . N Q I . N N K G S I I S . P
W M G N . N I I K Q T N I T V E N P I R F R A Q P Y A E I R S E I T K V P L Y L N
AGGAAATGCAGCATTAAATAACCAAGAGTAAATGGCTATTTTGAATGACCTGTATAGCAGAAAGTGAAGCAAAAGCAGGGGTATCTGGTGTGCATAGACTTGCAGGACACTTAAAGTAT 738
R K C S I N N Q E L N G Y F A M T C I A E S A K Q S R G I W C R I D F A G H L K Y
G N A A L I T K S . M A I L Q . P V . Q K V Q S K A G V S G V A . T L Q D T . S M
Q E M Q H . . P R V K W L F C N D L Y S R K C K A K Q G Y L V S H R L C R T L K V
GAATTGCTTAACAGCAATCTAACTGGGGGACGTATTCTCTTGACGCTTCCAGTAGCATTGAGTTGAAGTTAACTGCTTGAACAAATGCTACAAAACAATATCTGTATATTCTGTATATCTG 861
E L L N S N L T G G R I L L Q L P S S I R L K L T A . T N C Y K T I S V Y S V Y L
N C L T A I . L G D V F S C S F P V A F D . S . L L E Q I A T K Q Y L Y I L Y I C
. I A . Q Q S N W G T Y S L A A S Q . H S I E V N C L N K L L Q N N I C I F C I S
TATATAATATCCCGAATATCTGGGGATGCGCTATAAGTTAGCATGTTCTTGTTTTGCCTCATGAACAGATTAACCTGCTTTTGGGGCTGTTTTCTGACAGAGTGGGTTCCTGGTAAT 984
Y I I S P V I S G D A L . V S M F L F C P H E Q I K P A F G A V F L T E W V P G N
I . Y P Q . Y L G M R Y K L A C S C F A P M N R L N L L L G L F S . Q S G F L M
V Y N I P S N I W G C A I S . H V L V L P P . T D . T C F W G C F P D R V G S W .
GCTGAAGTCGCTTTTTTTTCCAGCAGAAAGATTAATAACACTTAACTGAATACTTTTCCATCATACAGTGTGCTGCTGATCGCATTGGCTGCAGAGAAAAACATGACAAAGATGCTCAGT 1107
A E A A F F F Q Q K D . N T L K L N T F P S Y T C H A D R I G C R E N M T K M L S
L K L P F F S S R K I K I H L N . I L F H H T R V M L I A L A A E K T . Q R C S V
C . S C L F F P A E R L K Y T . T E Y F S I I H V S C . S H W L Q R K H D K D A Q

TTCTATATAGTATTAAGAGTAATTTGTGCTCATCCCTTCTGTGCAGCAAGAGTAAAAGTGGTTTTGGTTGCTGGGCTTTAATTATTCTAACTCCAGGTGCATGAGTGAGATTTTTTTTCCA 1230
 F Y I V L R V I C A H P F L C S K S K T G F G C L G F N Y S N S R C M S E I F F P
 S I . Y . E . F V L I P S C A A R V K L V L V A W A L I I L T P G A . V R F F F Q
 F L Y S I K S N L C S S L P V Q Q E . N W F W L L G L . L F . L Q V H E . D F F S
 GATCTTGTGGAAGTGAAGAAGTCCCGGGAACAGGGCTGAAAATATCCTTAAGACTCTAGAAAATGACTGCAGTGTGGGTGAGAAGTGGCAGGCCCGTCTGTGCCGCGTGGGTAGGA 1353
 D L V G T . R T S R E Q G . K Y P . D S . K M T A V W V R T G R P R R V P R W V G
 I L L E L E E L P G N R A E N I L K T L R K . L Q C G . E L A G P V V C R A G . E
 R S C W N L K N F P G T G L K I S L R L L E N D C S V G E N W Q A P S C A A L G R
 GGCTGAAAGTGCAGGCTGAGCGTCTCGGTCGGCGGCGGTTTCGAGCTCCTTTGCAGCCTCTCATGAATGAAGTCCAGAGTTTTTTTATGAATGGGCGTCCGTCCTTGAATGAACACCAGTG 1476
 G . K C R L S V L G R R R F E L L C S L S . M N F Q S F F M N G R S V L E . T P V
 A E S A G . A C S V G G G S S S F A A S H E . T S R V F L . M G A P S L N E H Q C
 R L K V Q A E R A R S A A V R A P L Q P L M N E L P E F F Y E W A L R P . M N T S
 CGCCGCGCTGATTGGCCCCGGGATAAAGCCACGCCACTTTCCCCCAGCGTGTGCGATTGTTGGAGCGCCCCAGCCAGGGCTTATTACGAAGTCAATGAAACAAAGGAGCGGG 1599
 R R A L I G P G G . S H A H F P P S V L R L V D A P Q P R A S F T K S M K Q R S G
 A A L . L A P A D K A T P T F P P A C C D W W T R P S P G L H S R S Q . N K G A G
 A P R S D W P R R I K P R P L S P Q R A A I G G R A P A Q G F I H E V N E T K E R
 GAGCCGAGCTCGTGAAGCGGGCATGCCAATGGGAGGCCGCCCGGGGCTCAACGGCCTGAACGGCCGCGGGGGGGCTGCGTGGGGCTCCGCATGGGCCGGGAGTCCGCCACCGGGT 1722
 E P S S . R R R C A N G R P P P G A Q R P E R P R G R G C V G L R M G R E S P P G
 S R A R E G G D A P M G G R P P R A L N G L N G R A G G A A W G S A W A G S R H R V
 G A E L V K A A M R Q W E A A P G R S T A . T A A R A G L R G A P H G P G V A T G
 GAGGCGGGGCGGGAGGGTCTGGGGCTCGTCTGGGACCGGGGAGGGCTCATTCTCAGGGTCGGGAGTTCTGTCTACTCGTTACTGAAGTCTAGGGGCGGCTCGGCGCTCCCCTGGGGT 1845
 E A R G R R V G A R P G T G E G S F S G S G S S V S L A Y . S S R G G S A L P W G
 R R G A E G S G L V L G P G R A H S Q G R G V L F S H S L T E V L G A A R R S P G V
 . G A G P E G R G S W D R G G L I L R V G E F L C L T R L L K F . G R L G A P L G
 CCCGGTCTCCAGCCCTATTCTTACCCGGTGGGGGTGCTCCGCTTGTAGCCCCATCCTCGGCAGGTGCCGCGCCGCTGGTAGCCTTCTCTGTGCCTCCCCAGGTGCGCTCCGAG 1968
 P G S S S P L F L Y P V G G A P L V A . P I L G R C R P A G S L P L C L P Q V R S E
 P V L P A L Y S S T R W G V L R L . P P S S A G A G P L V A F L C A S P R C A P R
 S R F F Q P S I P L P G G G C S A C S P H P R Q V P A R W . P S S V P P P G A L R
 AGCCGGGAGCTCCGCGCCCGGAGGTAGCCGCCGGGGCGGAGGATGAAGGGGTACGTGGACAGCAGGTGCCGTTACTTTCCCCAGGTGAGTGGTGGCTTACCGGGCGTGGGAGGGGGA 2091
 S R E L R G R G G S A R A A G . R G T W T S R C R S L S P R . V V A L P G V G G G
 A G S S A A A E V A P G R Q D E G V R G P A G A V H L F P P G E W W L Y R A W E G D
 E P G A P R P R R . R P G G R M K G Y V D Q Q V P F T F P Q V S G G S T G R G R G
 TCAGGGCCGGACAGCCTCCGTTCCGGGCTCGGCTGATCCGCTCTTCCCGCAGCCGCCCGGGCCATCGGGGCTCCGGGTGCCGACTCCGACGGGCGTGC AAAAGGCCATGGAGCC 2214
 S G P G Q P P F G A S R . S A L P A A A A G P C G L R V P D S D G R C K E P H G A
 Q G R D S L R S G P R A D P L F P Q P P P G H A G S G C R T P T G A A K S P M E P
 I R A G T A S V R G L A L I R S S R S R R R A M R A P G A G L R R A L Q R A P W S

Exon 2
 Deduced Amino Acid Sequence Corresponding to the Open Reading Frame Containing Exons 3 and 4
 Putative Start Codon
 Putative Amino Acids Encoded by Exon 3

GGGCTTGCTGCCGGCGCAGGATTCCGAAGGTAAGGGTGGGGGGGGGCTGTAGGGGGAGGCTGGGGAAGGCTCTGGGCAGGGTCCAGGTCCTCCCCCTGTGCCTCCGCCCCACCTCCCCGTC 2337
G L A A G A G F G R . G W G G G C R G R L G K A L G R V Q V L P L C L R P H L P V
G L L P A Q D S E G K G G G G A V G G G W G R L W A G S R S S P C A S A P T S P S
R A C C R R R I R K V R V G G G L . G E A G E G S G Q G P G P P P V P P P P P R
CCGCCCGCATCTCCCCGCTGGAGCATCCCGCTGTCCTCCCCAGCGCTCCACTGAAGAAACCTCTCGCTTTTCGTGCAGACCTCTCCAGGACTTGAGCCAGTTCAGGAGACTTGCGTGAC 2460
P P G I S P L E H P A V P P Q R S T E E T S R F S C R P L P G L E P V P G D L A D
R P A S P R W S I P L S L P S A P L K K P L A F R A D L F Q D L S Q F Q E T W L T
P A R H L P A G A S R C P S P A L H . R N L S L F V Q T S S R T . A S S R R L G .
AGAAGTAGGCTCCTCCCCCTTCTCTGTTTGGTTTCTTCGCTTTTCGCTTCTTTGTCTTTTGGTGTGTTACTGTTCTTTGCTTTGTTTGGCCAGCTCAAACATTGAGGATTTTTTTC 2583
R R . A P S P F S V L V S S S F A S L S F W C L Y C S F A L F A Q L Q T L R I F F
E G R L L P L S L F W F L R L S L L C L F G V C T V P L L C L P S S K H . G F F S
Q K V G S F P F L C F G F F V F R F F V F L V F L C F V C P A P N I E D F F
AGGCTTGCCAACCTCTCGCTGTGTAATGAAGAAGCATGAATCACACCATTGNCTGGGTATTTGCTTATGAATGAACTTGCCCGCCAGCCGTTTTCTGGTACAATTTTTGTTTTCAGCCT 2706
R L A N L C A V . M K K H E S H H ? L G I C L . M E L G R Q P V F L V Q F L F Q P
G L P T S A L C K . R S M N H T I ? W V F A Y E W N L A A S P F S W Y N F C F S L
Q A C Q P L R C V N E E A . I T P L ? G Y L L M N G T W P P A R F P G T I F V S A
CTTCTGTAACCGTTTTGATTTTATGCTAATTTGTCGCTGATCTTTGTTAGTTACCAAACATTATATTTCTTAAGCAAACAATGAAGTGTATTTTAACTGGGAGACGAGGTGCTTTTC 2829
L L V T G L I L C . F V A D L C . L P K H L Y F L K Q T M N C Y F N L G D E V L F
F L . P V . F Y A N L S L I F V S Y Q N I Y I S L S K Q . T V I L I W E T R C F S
S S C N R F D F M L I C R . S L L V T K T F I F P . A N N E L L F . S G R R G A F
CTAACGTTTTTTCTTTGGAGCATAATTCTCTGAAGCACCCTGCTGGAGTTCATCTTTCTACATTAGCATCTAAACTGTTTGCCTCTTGACTTCAAGGAAGAGAAAACCTCAAGTGTGG 2952
L T L F S F G R . F S A S T V L E F I F L H . H L N C F A S . L Q G R E N L K C W
. R F F L L D D N S L Q A P S W S S S F Y I S I . T V L P L D F K E E K T S S V G
P N A F F W T I I L C K H R P G V H L S T L A S K L F C L L T S R K R K P Q V L
GGAGTGCCGTTAATCCAAGTGAAAATAGTCTAAGCAATGGAGTGAACCTTTGTTCTAATTTCACTACTCCAATGATCTATTTTGTGAGGAACTGAAAAGGTTTCGTTAAGTACAGACTCC 3075
G V P L I P S E N S S K Q W S E L C S N F S T P N D L F C . G T E K V S L S T D S
E C R . F Q V K I V L S N G V N F V L I S V L Q M I Y F V E E L K R F R . V Q T P
G S A V N S K . K . F . A M E . T L F . F Q Y S K . S I L L R N . K G F V K Y R L
TGATAAATAGGAGACCTTACTGTGATATCAGTTCTTAATAGTAAATAGTTGCTAAGTAAAGAAATGAAATACCTTATGTTCTTAAATTTGGAGTGGGTAGAAATCAGTGAAGATTAA 3198
. . I G E H L L . Y Q F L I V K . L L S K E M K Y L M F L K I W S G . N S V R I .
D K . E S T Y C D I S S . . L N S C . V K K . N T L C S L K F G V G R I Q . G F K
L I N R R A L T V I S V L N S . I V A K . R N E I P Y V P . N L E W V E F S E D L
GGGTACAATAGGAATTAAGTGAAGCTGTTCTGGGTTGAAACACCGCTGATTGAAGCAGGAATGTAGGTATCACTGGTCTGTTGGATCTGATGAAGGTGCACACTTGCTGGGACTGAT 3321
G Y N R N . L R L F W G . N T A D L K Q E C R Y H W S S V G S D E G A H L L G T D
G T I G I N . G C S G V E T P L I . S R N V G I T G R L L D L M K V H T C L G L I
R V Q . E L T E A V L G L K H R . F E A G M . V S L V V C W I . . R C T L A W D .

Exon 2
 Deduced Amino Acid Sequence Corresponding to the Open Reading Frame Containing Exons 3 and 4
 Putative Amino Acids Encoded by Exon 3
 Putative Amino Acids Encoded by Exon 4

TTCTCCATTGAGATGTTTATATCACAGTCAGATTGTGCCAGTCTTTGACTTTGAGTTTTCTGATGCTTGATAATATTTTTTTCAGATATTAATCCAAAAGAAATTTGTGACCATCTGTGTAATC 3444
F S I E M F I S Q S D C A S L S T L S F L M L D N I F Q I L I Q R I C D H L C E F
S P L R C L Y H S Q I V P V F R L . V F . C L I I F F R Y . S K E F V T I C V N S
F L H . D V Y I T V R L C Q S F D F E F S D A . . Y F S D I N P K N L . P S V . I
ATCCCTGCTCTGCTACTGGCAGACGTTCCATAGTCAGTGCCTCAGGTGGAAGTGAGGTTGCATTGAGGTCTCATAGCAGTGGATGGAAGAAAGGGAATTTGTCATTAATGTGTCTCATTATGTAT 3567
I P A L L L A D V P . S V P Q V E V R L H . G L I A V D G K R E L S L M C L I M Y
S L L C Y W Q T F H S Q C L R W K . G C I E V S . Q W M E K G N C H . C V S L C I
H P C S A T G R R S I V S A S G G S E V A L R S H S S G W K K G I V I N V S H Y V
AATGCAAACCTTAGAGGAAAAGTAGCAGCACAGTGTATGATTCCAGCTGAGTTCCTTCTGCCTGGAGGAAGAACTCTTAATAGCCCTGAGGATGGAAGGAACTTTGGGTTGTGCTAACAGC 3690
N A N L R G K V A A Q C Y D S S . V P S A W R K N S L I A L R M E G T L G C A N S
M Q T L E E K . Q H S V M I P A E F L L P G G R T L . . P . G W K E L W V V L T A
. C K P . R K S S S T V L . F Q L S S L F C L E E E L F N S P E D G R N F G L C . Q
TCCGCTCTTTCATCCTATCAGATGAAATCCCCCTTCTCACTCTTATGAATACTCCGAGCTTTAAATGCCCTTCACTCCAATACTCAGGAACTTTGTTGCTTCTATGAATAC 3813
S A L C I L S D E I P P L S H S L . I L R S F K M P F T P I T S G T L L L S M N Y
P L F A S Y Q M K S P P F L T L Y E Y S V A L K C L S L Q . L Q E L C C F L . I T
L R S L H P I R . N P P P F S L F M N T P . L . N A F H S N N F R N F V A F Y E L
TCCTCAGGATGCACTCAGCAATGGTGTGCTGCTCTCAATCTTCTGCTGGGGCAGCTGAGTGGCGAAGCATCCAGCATCTTATCGAGATCCCACAGCTAAGGATGGGGCTCTGGGTTTCC 3936
S S R C T Q Q W C A V L S I F C W G Q L S G E A S Q H L I E I P Q L R M G L W V S
P H D A L S N G V L C S Q S S A G G S . V A K H P S I L S R S H S . G W G S G F P
L L T M H S A M V C C A L N L L L G A A E W R S I P A S Y R D P T A K D G A L G F
TATTACCTGATTTTATGCTGCAGTCACAAAAGCGCAGTGCCTGCAGTGCCTCATAAACGTAATGACTTTTCTTTTGTGATGGACATGCAGAGTATTCTTGAATTAATAATGAA 4059
Y Y L I F M L Q S Q K A Q C L Q C S V P H K T . . L F F L M D M Q S I L E I T N E
I T . F L C C S H K K R S A C S A Q C L I K R N D F S F . W T C R V F L K L L M N
L L P D F Y A A V T K S A V P A V L S A S . N V M T F L F D G H A E Y S . N Y . .
TCTTCGGAGCTGAGAAGACAAGATTTTAAATGGGACAAAATGAAATGAGAGGGTAGAGCCAGAAAATACCAGCCATTAACCATGGAAGGCTTCAGAGGTGGCTTTGTGCGGTGGGTTATTTCC 4182
S S E L R R Q D F N G D K M K . E G R A Q K Y Q P L K H G R L Q R W L C A V G Y F
L R S . E D K I L M G T K . N E R V E P R N T S H . N M E G F R G G F V R W V I S
I F G A E K T R F . W G Q N E M R G . S P E I P A I K T W K A S E V A L C G G L F
TGAGAGGTTACAGTAGAAGTGGAGAACAAAAATCTGTTGTAGGCTGTTTGTGCTGCTGTTAGCAGGCAGATCTTCCGTCCTTATCCTTGTACCAGATTAACCTCCAGGCTGAATG 4305
. E V T V E V E N K N L L . A V L L L S V S R Q I F R P L Y P C Y Q I K P P G . M
E R L Q . K W R T K I C C R L F C C L L A G R S S V P F I L V T R L N L Q A E C
L R G Y S R S G E Q K S V V G C F V A V C . Q A D L P S P L S L L P D . T S R L N
CAAGAGAATGCTTGCAGAAAATGCCCTGTGATTTAGAAAAGACTGCATGCTTTCAGAAAAGAGGAGGTTTTATATGGCTCATCCACATCCAGAATCTGTGAGAATGTGCTGATGTGACGC 4428
Q E K C L R E N A L . F R N D C M L S E K R K V L Y G S S T S R I C Q N V L M . R
K R N A C E K M P C D L E T T A C F Q K R G R F Y M A H P H P E S V R M C . C D A
A R E M L A R K C P V I . K R L H A F R K E E G F I W L I H I Q N L S E C A D V T

TGGCAGATGGTGACAGTGCAGAAGTGATCTTGAATGAAGCTGTGAGCAAAGCGACGCTCAATAAGCTGAGAAACCAACCTCGGCTCAGACGGTGTGATCGTATTATTTCTCCTAAGCCACT 4551
W Q M V T V Q K . S W N E A V S K A T L N K L R N P T S A Q T V . S Y Y F S . A T
G R W . Q C R S D L G M K L S A K R R S I S . E T Q P R L R R C D R I I S P K P L
L A D G D S A E V I L E . S C Q Q S D A Q . A E K P N L G S D G V I V L F L L S H
TTCTGAAGGTGGTTCACACAATGAATTGGGCAGGTTTGCCTTGTGCCCTCCCTTTGTTGCTTATGGTCCCGTTAGCCGTCACATGGGTTATTTAAGGAGCAGTGATTGCAGTGAGAAGA 4674
F . R W F T Q . I G Q V C F C S P P F V A Y G S R . P S H G L F . G A V I A V R R
S E G G S H N E L G R F A S V P L P L L L M V P V S R H M G Y F K E Q . L Q . E E
F L K V V H T M N W A G L L L F P S L C C L W F P L A V T W V I L R S S D C S E K
GTCAGCTAGTCCACTCTAATCCACTGCCATTGTAGCATGGGATTATCCCTGAGGTTTATTTTGGCATGTCTTGTATCTTGGCTTACATGACCGAGATTAGTAGGCTTCCATTCC 4797
A Q L V P L . S T A H C . H G I I P . G L F L A C L V Y L G F T . P R L V G F H S
L S . F H S N P L P I V S M G L F P E V Y F W H V L S I L A S H D R D . . A S I P
S S A S S T L I H C P L L A W D Y S L R F I F G M S C L S W L H M T E I S R L P F
TCATCTCCCTTACTCTGACTGCATTGCTTTAGAGCTAAGTGCAGACTTCTGTTACGCCCAATATCCAGAGCAAAGCAAGGTCTGCAGTGAACCAGCAGTATATTTGTGAATGGTTTTGT 4920
S S P F T L T A L L . S . V Q T S V H A A I S R A K Q G L Q . T S S I F C E W F C
H L P L L . L H C F R A K C R L L F T P Q Y P E Q S K V C S E P A V Y F V N G F V
L I S L Y S D C I A L E L S A D F C S R R N I Q S K A R S A V N Q Q Y I L . M V L
GACTTAAGCTGCAAGCTCTCCCTCCCTGCAGTTAAAGGGACATAAAAGTAATAGCTGGAAGTTATAAAATATTATTGCAACNCGCGTGTGCTGTGGTGAAGGTTCTCCCCGAGATGATTCT 5043
D L S C K L S L P C S . R D I K V I A G S Y K I L L Q ? A C A V V K V L P E M I P
T . A A S S P F P A V K G T . K . . L E V I K Y Y C N ? R V L W . R F S P R . F L
. L K L Q A L P S L Q L K G H K S N S W K L . N I I A T R V C C G E G S P R D D S
CCCAGCTGTGTGAGTGGGCTGTGGGATGTGGGCTCAGCTGCACCTGAGCTGAGGGAAGCAGAGCACCCCATCACCAAGGATGGGGCTCAGTAGTGTGTGCTTGGAAACACGGGAAAG 5166
P S C C A V R A V G C G L S C T . A E G S R A P P S P R M G L S S V C A W N T G K
P A A V Q . G L W D V G S A A P E L R E A E H P H H Q G W G S V V F V L G T R E R
S Q L L C S E G C G M W A Q L H L S . G K Q S T P I T K D G A Q . C L C L E H G K
GATTATCCCTGTATGGATGTGCAAGGTTGAAGTTAGGAATGCTGCTTGGCAAATGGGAGATGCAGGAAGTGAAGGAGGACATGTGTGGCTGACGGCTGTGAAATCACAAGCTGGTAACT 5289
D Y S L Y G C A R V K L G M L L L A N G R C R K . R R T C V A D G C E I T K L V T
I I P C M D V Q G . S . E C C S W Q M G D A G S E G G H V W L T A V K S Q S W . L
G L F P V W M C K G E V R N A A L G K W E M Q E V K E D M C G . R L . N H K A G N
ACAACAACAAGCAGCACCAGTAAGATATCCAGCATTCTGGTGTCAAACCTGTACAGCAGCACTCTGGACTATGAAAGCATTGATCTTTTGAATGTTAATCTGCCAAAAACACACTTGAG 5412
T T N K A A P V R Y P S I L V S N L Y S S T L D Y E S I R S F E C . S A K N T L E
Q Q T K Q H Q . D I P A F W C Q T C T A A L W T M K A F D L L N V N L P K T H L S
Y N K Q S S T S K I S Q H S G V K P V Q Q H S G L . K H S I F . M L I C Q K H T .
CACCTTGACGTCAAACTCCTTCTCTATTGTGCTGCTGACTTTGTGCCATGAGGCTCTCGTTTTGGCTGAGGGAGACATTAAGTCGGACTTTGTGCTGTTAGTGAGATGCTGAAGTTGTTC A 5535
H L A V K L S P L L C C . L C A M R L S F V L A E G D I K S D F V L L V R C . S C S
T L Q S N S L P L Y C A A D F V P . G S S F W L R E T L S R T L C C . . D A E V V Q
A P C S Q T P S S I V L L T L C H E A L R F G . G R H . V G L C A V S E M L K L F

GCAATATCTTCTTCCAAGTATCATTATTTCTCCTAAATGTTGCTTCTGCTGTCTACAAGTGGTCTCCTATCCTGCTAGCAGCTATCCCCTGCTGCTGGAGCTGCTGAAAGGATACTTC 5658
A I S S S K Y H Y F P P K C C L P V C L Q V V S Y P A S S S Y P L L L E L L L K G Y F
Q Y L L P S I I I F L L N V V F L S V Y K W S P I L L A A I P C C W S C . K D T S
S N I F F Q V S L F S S . M L S S C L S T S G L L S C . Q L S P A A G A A E R I L
CTATACAGAGATGATAGTGTTCCTGCTCCTCCTGCTGCTTTGGCACAGCCTTTCCAAATCCACCAGCTGGTCCAGAGCCTTCTTACTGAATGCTTAAAAAAGATCTACCAATGATGCT 5781
L Y R D V G V S C S P R C F G T A F P N P P A G Q S V S F T E C L K K D L T N D A
Y T E M . V F P A L L A A L A Q P F Q I H Q L V R A F P L L N A . K K I S P M M L
P I Q R C R C F L L S S L L W H S L S K S T S W S E R F L Y . M L K K R S H Q . C
GGTTGGGAGAGTGAACCTGGAGCTAATAAACGTCACCCCTTTTGTCTCCAAGGAAGGGCGGGATGCTGAGCCACAGCTCTCGGGCAGCAGAGCTGGAGTCATGAATCCCAAAGGGAAGCA 5904
G W E E . T W S . . T S P F L L P R K G R D A E P Q L C G Q Q S W S H E S Q R E A
V G K S E P G A N K R H P F C S Q G R G G M L S H S S A G S R A G V M N P K G K Q
W L G R V N L E L I N V T L F A P K E G A G C . A T A L R A A E L E S . I P K G S
GCTCATCGGCACCAAAGTGCCTCCCCACAGAAGGCTCAGGTCTTTCCAGCGTAGAATCATTCTACCTTACTGTTAATTGCACCATACAAAGTGCCAAATCTAATTCTCCTCTCTCTCTCT 6027
A H R H Q S A S P Q K A Q V F P A . N H S Y L T V N C T I Q S A K S N S P L S L S
L I G T K V P P H R R L R S F Q R R I I P T L L L I A P Y K V P N L I L L S L S L
S S S A P K C L P T E G S G L S S V E S F L P Y C . L H H T K C Q I . F S S L S L
CTCTTCTCCCTCCCGGAGTCAAGTTCAGACAGCGATGAGCAATTTGTGCTGACTTCCATTAGAAAACCGTAAGTGGCGGGGGCGGCTGTGCACCTCAGGGGAGAAGGGCTGGAT 6150
L F S P P P G S Q V P D S D E Q F V P D F H S E N R K W R G A A V H L R G E G L D
S S P L P P A V R F Q T A M S N L C L T S I Q K T V S G A G R L C T S G E K G W I
S L L P S P R Q S G S R Q R . A I C A . L P F R K P . V A R G G C A P Q G R R A G
ATGTTTACAGGATCCATTTACCCCAAGGGCTCTGCAGATGTGGCACAGATATGGAGGCTGAAGTGTCTGCTCAGCCCTGCTTTGTTGTGACAGTGACAACGAAGCCCTGGTGTCCATT 6273
M F T G S I Y P Q R A L Q M W H R Y G G . S A L L S P A L L . Q . Q R S P G V P I
C L Q D P F T P K G L C R C G T D M E A E V L C S A L L C C D S D N E A L V F P F
Y V Y R I H L P P K G S A D V A Q I W R L K C S A Q P C F V V T V T T K P W C S H
CATAAATGGAGGATATCTGCTGCTCGGAGCTGATCAAACCTTATGCTGAGCTGATTGAGTGGGTGTTTTCAGGTGGTGCACATGTGAATGGTCCAGGCATCCTGTGAAGAGCTTACAGAG 6396
H T M E G Y L R A R S . S N L M L S . F S G C V S G G A H V N G Q A S C E E L T E
I Q W R D I C V L G A D Q T L C . A D S V G V F Q V V H M . M V R H P V K S L Q S
S Y N G G I S A C S E L I K P Y A E L I Q W V C F R W C T C E W S G I L . R A Y R
TGCCTAATATGATAACCAAGGTAGAATTTGTGCTGCTCTCATCCAGAGTGAAGGAGTGGGCTATGTGTGAAATCTGGGTGCTGTACATCCAGGTGGGGATGCTTTGGGGCTCTGC 6519
C V N I T K G R I C A V S H P R V K G V G L C V E F W V L Y I P G G D A L G L C
A L I . . P K V E F V L S L I P E . K E L G Y V W N S G C C T S Q V G M L W G S A
V R . Y D N Q R . N L C C L S S Q S E R S W A M C G I L G A V H P R W G C F G A L
TCACCACATGAGCTGCTGCTGGGATCACATCCACTGCCATTAGGAAGTGAAGATACGCTGAAAATAGCCCTGTTTGACCAGAGTGTGCAAAGCAGAGCATTGATCCATCCTGC 6642
S P T . A R A A G I T S H C H . E L K I R A E N S P V C T R V C K A Q H L H P S C
H P H E L V L L G S H P T A I R N . R Y V L K I A L F A P E C A K H S I C I H P A
L T H M S S C C W D H I P L P L G T E D T C . K . P C L H Q S V Q S T A F A S I L



Exon 5

AGAGCAGTGAGCACTGTGCAGGTCACATGGGGCTGGGAGCTTCTGCCATGGCTGTGCAGCAATTCCTTCTCTTTCTGCCAAATGCCAACGGCACAGGGTCCATCGTCTGGGC 6765
R A V S T V Q V P H G A G S F C P W L C S N S F S F P A Q M P T A Q R S L H R L G
E Q . A L C R S H M G L G A S A H G C A A I P S L F L P K C Q R H S G R S I V W A
Q S S E H C A G P T W G W E L L P M A V Q Q F L L F S C P N A N G T A V A P S S G
AGCTCTGGGAAGTTCATTAGAGCCATTGCGGTGCGGTGTGTATCCCACCTCTGTCTCTCTGCTTGGGGCAGTTTTCCCTATGAAAACAAAACGAAATGCTCCTCAAGCGGTCT 6888
S S G K F I Q S H C G R L C C Y P T S V P L C L G Q F S L . K P K R N A P P S G S
A L G S S F R A I A V G C A V I P P L F L S A W G S F P Y E N Q N E M L L Q A V L
Q L W E V H S E P L R S A V L L S H L C S S L L G A V F P M K T K T K C S S K R F
CTCCTCCCTCCTCCCGACTTGGCTCTGGGTGTGCTCTCTGAGTCATGCAGCTTGTTCCTCTCTTTGACCTAGAACTAAGAGGAGGGCTGCCATGTTTCGCTCTGAATGAATGAGCC 7011
L L P P S R L G S G C A L . V M Q L V S S S L . P R T K R R A V P C F A L N E . A
S S L L P D L A L G V L S E S C S L F L P L F D L E L R G G L S H V S L . M N E P
S P P S F P T W L W V C S L S H A A C F F L S L T . N . E E G C P M F R S E . M S
AGTTGTAATGCGATCGCTGTTCCAGCTAGCAGGGGCCAGAAAGGAGTGGGCTGCTCCAGCACATTTTACTGCTTGAGGTTTCAATGGAGCTCAGTGTACTGAGCAGAGCATTTCTATTGGT 7134
S C N A I A C S S . Q G P E R S G L L P A H F T A . G F N G A Q C T E Q S I S I G
V V M R S P V P A S R G Q K G V G C S Q H I L L L E V S M E L S V L S R A F L L V
Q L . C D R L F Q L A G A R K E W A A P S T F Y C L R F Q W S S V Y . A E H F Y W
CTTTATTTTCTGCGAGGCTCAGCCAACATCAGTGTTTTTAGATGGGTCTGTACATCCCTTTATGCTGTACATTGGGAAGAGCAGCTCAGGGCTGCCATGGGAGCCCCCTATCCCCAG 7257
L L F F L Q A Q P T H Q C F . M G S V H P F M L Y I G K S S S G L P M G A P H P Q
F Y F S C R L S Q H I S V F R W V L Y I P L C C T L G R A A Q G C P W E P L I P R
S F I F P A G S A N T S V F L D G F C T S L Y A V H W E E Q L R A A H G S P S S P
GGAGGAGCAGGAGCCAAAAGCAGTTTGGGCTCTGCTCTGATGGTTGATTTATTTACTCTGATTCGTGAGCTGCTGGATCCCTCTGGGTGCTGGCATGATGCAGAGGGTGAAGGGAGG 7380
G G A G A K S S L R L C S L M V V F I Y S D S V S C L D P S G C W H D A E G E G R
E E Q E P K A V . G S A L . W L Y L F T L I L S A A W I P L G A G M M Q R V K G G
G R S R S Q K Q F E A L L S D G C I Y L L . F C Q L L G S L W V L A . C R G . R E
CTTCTGAGTGTGCTGAATCCCAAGGCAGGCAACAGAGCCAGGCTGGTGGGTTTTTTTTTCCCCCATCTTGCTGAGATCAAATCTGGAGCTGGTGGGGGAAGTGTGCAGTAAGAAG 7503
L S W S V L N P K A G K Q S Q A G G F F F S P I L L R S N S G A G W G K C C S K K
F P G V C . I P R Q A N R A R L V G F F F P P S C . D Q I L E L G G G S A A V R S
A F L E C A E S Q G R Q T E P G W W V F F F P H L A E I K F W S W V G E V L Q . E
TGTCAGAACGTCATGTACTGAGAGAAAGAACAGCAGCCCCCTCCAGCCCTTGGCCAGCAGCTTGGTGGGGCACTTAGGAATTGCCAGGCTTGCACTGAGGCACAGCTGGTCTCCTGTGG 7626
C Q N V M Y . E K E Q Q P L P Q P S C Q L A G A L R N C Q A C S E A Q L V L L W
V R T S C T E R K N S S P F P S P L A S S L L G H L G I A R L A V R H S W C S C G
V S E R H V L R E R T A A P S P A L L P A A C W G T . E L P G L Q . G T A G A P V
CACGGATTAGTGCAGAAACAGCCAGCAAAGTGAGTTATTCGCGGCTCAGGATGTGATCTCCACCTTCAAACCTGTAATCACAGCTCTGGACATCCCACTCCTTGAAGTAGGGTTGTGTC 7749
H G L V Q K Q T Q Q S E L F A A Q D V I F H P S N C N H S S G H P T P C K . G C V
T D . C R N R P S K V S Y S R L R M . S S T L Q T V I T A L D I P L L V S R V V S
A R I S A E T D P A K . V I R G S G C D L P P F K L . S Q L W T S H S L . V G L C

AGCTCAAAGAGGTGGGAGAACTGGCGAAGGGACTGCAGTGAGCCTTGGGAAGGGCTCAGCCCTCGTTTGGGAGACCGACCTCCTCCTTGTGTACAAACAGAGCTGGGTACTTCTTTGGG 7872
S S K R W E E T G E G T A V S L G K G S A L V W E T D L S L V C T N R A G Y F F G
A Q R G G R K L A K G L Q . A L G R A Q P S F G R P T S P L F V Q T E L G T S L G
Q L K E V G G N W R R D C S E P W E G L S P R L G D R P L P C L Y K Q S W V L L W
GTACTCTTCTCCTCCCTAAGTGCATAGCCTCCTACCAGCAGCAGGGTCCATAGGCAGAAGATGCTCACCACATCCCTAAGCCAGGGTGAAGTATTCAGAGGAATCGGTGAAGGAAGTTG 7995
V T L P S S L S A I A S Y Q Q Q G P . A E D A H H I P K P G . D F Q R N R . R K L
L L F L P P . V P . P P T S S R V H R Q K M L T T S L S Q G E I S R G I G E G S C
G Y S S F L P K C H S L L P A A G S I G R R C S P H P . A R V R F P E E S V K E V
TGTTCCCTGCTGGATTTAATTAAGCACAGAAGTAACGCTACCCACATCCTGACACTGTGAACCTGCAACCCACATTTGTTATTCATGAAACCTTAACCAGCCCAGAAATAGTCTCCTATTAT 8118
C S L L D L I K H K K . R Y P H P D T V N C N P H L L F M K P . P A Q K . S S Y Y
V P C W I . L S T R S N A T H I L T L . T A T H I C Y S . N L N Q P R N S P P I I
V F P A G F N . A Q E V T L P T S . H C E L Q P T F V I H E T L T S P E I V L L L
AGGCTAGGAGGGAGGCTGCAGAGGAGACCTGCAAGCTGCTTCCAGCTTGGAGGTGGGTTTGGTGAACACAGGGGCTGAAGTCAGGTTCTGCTGTGTGCTCTGTGTGAGCAGTATGCAGC 8241
R L G G R L Q R G P A S C S S S L E V G L V K T G A . S Q V L L L C A L C E Q . C S
G . E G G C R E D L Q A A L P A W R W V W . K Q G P E V R F C C V L C V S S D A A
. A R R E A A E R T C K L L F Q L G G G F G E N R G L K S G S A V C S V . A V M Q
TCTTCTAGAACAGGACCTCCAGTTCTGGGTTGGTTGTGCTTGGAGATGCTTCTGTATGAGAGTTAGTCTGGTGACTGCCTGTAATAGGATGCAATGTATGTAATAGGATGCAATCTATAA 8364
S S R T G P P V L G W L C F W R C A S L M Y E S . S G D C L . . D A M Y V I G C N I .
L L E Q D L Q F W V G C A S G D A S C M R V S L V T A C N R M Q C M . . D A I Y N
L F . N R T S S S G L V V L L E M L L V . E L V W . L P V I G C N V C N R M Q S I
TAGTATTTCCAGGAAAGCCCTCAGCTCATTGTCTACAGAAGCCTCACACTTATTTTCATGCAGGACAACCATACAATCAGGAAATGCACAGTGTGGCATTGTGTTATTGCATTATAACT 8487
. Y F P R E S P Q L I V Y R S L T L I F M Q D K P Y N Q E M H S V A L C Y C I I T
S I F P G K A L S S L S T E A S H L F S C R T N H T I R K C T V W H C V I A L . L
I V F S Q G K P S A H C L Q K P H T Y F H A G Q T I Q S G N A Q C G I V L L L H Y N
TAATGCAGTGTACGTGGTAATGCTACCCATAAGACATGCTTAAACAAGTATGCCAACAGAAGAGAGTTTTGCCTGATGATGCAAAACCAAGCTTTGCTTTCATGTGTTGAAAAATAGGATCC 8610
. C S V T W . C Y P . D M L N K Y A N R R E F C L M M Q T K A L L S C V E K . D S
N A V L R G N A T H K T C L T S M P T E E S F A . . C K P K L C F H V L K N R I P
L M Q C Y V V M L P I R H A . Q V C Q Q K R V L P D D A N Q S F A F M C . K I G F
TGCATGCAACCCCTCGGGAGTTTCTCCATATATGAGAGAAGTCTCAGCTTGATTTCCACCTCCTCCCTCAGTAGCTTTTCATAGTCTCCTGCAAAGATTAAGAAAGAACCCAGAGTCC 8733
C M Q N P L G S F S I Y E R S L S L I F P P L P L S S F S . S S C K D . K G T T E S
A C C P L S G V S F S P Y M R E V S A . F S H L S P S V A F H S P P A K I K K E P Q S P
L H A T P R E F L H I . E K S Q L D F P T S P P Q . L F I V L L Q R L K R N H R V
CTCCTCGGACCCACGCTGCTCCTGCAACCAAGCAGCCATTCCAGTACCGAATGGAGAGCAGTGCCTTTATGCCAGGTAACACAGGGAGGGACGGGGCTGTTGGGAAATGGGTGCTGCT 8856
L L G P H A V L Q P Q A A I P V P Q W R A V P L C Q V T Q G G T G A V G K L G A A
S S D P T L S C N H K Q P F Q Y R N G E Q C L Y A R . H R E G R G L L G N W V L L
P P R T P R C P A T T S S H S S T A M E S S A F M P G N T G R D G G C W E I G C C

Exon 6

GTGTTATCTGATGAGCTCTTATTGCCACAACCTGACCTGGGGGAGAGCAGTGTGGTATGGATCTGCAGCCACGGATCACGCAGGGGTTGGTATTTCCAATGCATGGCATAGCTCAGAGCCATG 8979
V L S D E L L L P Q P D L G E S S V V W I C S H G S R R G W Y F Q C M A . L R A M
C Y L M S S Y C H N L T W G R A V W Y G S A A T D H A G V G I S N A W H S S E P W
C V I . . A L I A T T . P G G E Q C G M D L Q P R I T Q G L V F P M H G I A Q S H
GGGAGATCAGGGCCAGCATCTTTTTCTCCCTAGTGCTGATGAGAGCAATACTGAGTGCATAATAACCCATAGGATTGTAACGCTCCAATATCTGTTGACTCCCCAGTGCCATGATC 9102
G E I R A S I F F S S L V L M R A I L S A . . P I G L . R S N I C L T P P V P M I
G R S G P A S F F P P . C . . E Q Y . V H N N P . D C N A P I S V . L P Q C L . S
G G D Q G Q H L F F L P S A D E S N T E C I I T H R I V T L Q Y L F D S P S A Y D
AACCCAGACCGTTGGGATCAAATCCCTGACCCAGGAACCCCAATCCAGTACCCTCCAGCAGTCCCCAGACAGGACAGGAACCTCCTGACCCCTCAGGGCCACCCACCCACATCCA 9225
N P D R L G S N P L T Q E P Q S S H R S S S S P D R T G T S . P L Q G H P T P H P
T Q T G W D Q I P . P R N P N P V T A P A V P Q T G Q E L P D P F R A T P P H I Q
Q P R P V G I K S P D P G T P I Q S P L Q Q F P R Q D R N F L T P S G P P H P T S
GCTGTGGATACCTCATTGAAAAAGGTGAGCTGTGGGTGGAGGTGGCAGCCAGCAGTGTGGTCTGTCTGATTGGGTTTGGGAAGGACAGGGGACTACACTAGCAGAGGGTGGCAGCATGGCA 9348
A V D T S L K T G E L W V E V A A S S W V C L I G F G K D R G L H . Q R G G S M A
L W I P H . K Q V S C G W R W Q P A V G S V . L G L G R T G D Y T S R G V A A W Q
S C G Y L I E N R . A V G G G G S Q Q L G L S D W V W E G Q G T T L A E G W Q H G
GGGAGGGGACTGTGCCCTCTGCTCTGCCCTTATGGGACCCAGCACAACAAGATGTGGGTGGGTCCAGAGGAGAACACAAGGATTATCAGAGGCTGCAGAACCTCTGCTTTGGAGAAA 9471
G R G L C P S A L P L W D P Q H N K D V G W V Q R R T Q G L S E A A E P L L W R K
G G D C A P L L C P Y G T P S T T K M W G G S R G E H K D Y Q R L Q N L C F G E R
R E G T V P L C S A L M G P P A Q Q R C G V G P E E N T R I I R G C R T S A L E K
GTTGAGGAGAAGAGAAGGCTCTGGGGAGGCCCTATTGCAGCCTCCTTGCTTGAGAGGAGCTTATAAACAGGAGGAGAATGCCTTTAACTAAAAGGGGAGATGTAGGTGAGATGCA 9594
V E G E E K A L G E A S L Q P P C A . E E L I N R R E N A F K L K E G R C R . D A
L R E K R R L W G R P H C S L L V L E R S L . T G G R M P L N . K R G D V G E M Q
G . G R R E G S G G G L I A A S L C L R G A Y K Q E G E C L . T K R G E M . V R C
AGGAGGAAATCTTTACCAGAGGGCAGTAAGGCTGTGGCACTGCTGCCAGAGCTGGGGTCCCAACCCCTGAAGGTGCCATGGCATGGATGGACCCTGGGCTGCCTGAGCTGTGGAGG 9717
R R K F F T Q R A V R L W H C C P E L G C Q P L K V P M A M D G P W A A . A A G G
G G N S L P R G Q . G C G T A A Q S W G A N P . R C P W P W M D P G L P E L L E G
K E E I L Y P E G S K A V A L L P R A G V P T P E G A H G H G W T L G C L S C W R
CACTGAGCCATGGCAGGGGAGTGAACACAGATTTGTTAGGGCCCTTCTAATACAGCTGTTCCATGACCCAGACCAAAGTGNTCAGTTTNTAGCAATCCCCCCCCACCTGAAGCACCTT 9840
H . A H G R G S G T R F V . G P F . Y S C S M T Q T K V ? S ? . Q S P P T L K H L
T E P M A G G V E P D L F R A P S N T A V P . P R P K ? S V ? S N P P P P . S T L
A L S P W Q G E W N Q I C L G P L L I Q L F H D P D Q S ? Q F ? A I P P H P E A P
GAATCCACAGCTGCTGCTGGGGTTGAAGGGAAGTCTGACATGAGGCTCATCAATTGTTGTGAATGATGGGCTGTTGTTGTGATTGCAAATACCCACTGTAGGTTCTGCTCCCTGCCCA 9963
E S T A A A W G L K G S L T . G S S I V V N D G L C C C H C K Y P L . V L L P A P
N P Q L L L G V . R E V . H E A H Q L L . M M G C V V V I A N T H C R F C S L P H
. I H S C C L G F E G K S D M R L I N C C E . W A V L L S L Q I P T V G S A P C P

Exon 7

TCTCTGTGTGCAGCTCCTCCAGCCCTCACAAACAGCAGCACGAGGCCCTTTGTCCCCTTGACCAGCCCCACACCAGCTCCCAAGTCCAGAAAGCATCCCAAGTCTCTCTTCTTGCAGCT 10086
 S L C A A P P A L T T S S T R P F V P L T S P T P A P K S R K A S Q V L S S L A A
 L C V Q L L Q P S Q Q A A R G P L S P . P A P H Q L P S P G K H P K S S L L L Q L
 I S V C S S S S P H N K Q H E A L C P L D Q P H T S S Q V Q E S I P S P L F S C S
CTGCCTACCAGCCACCTGTGGAGATGTGCCACTCCTCCCTCCCTCCAGGGCAGTGGCCAGGATGCTCCCATCGCTACCAGCGCCCATGTCCGAGCAATGCCTTCAGTTCCCCCGCAGG 10209
 L P T S H L W R C A T P S L P P R A W P R M L P S P T S A P C P S N A F S S P R R
 C L P A T C G D V P L L P S L P G H G P G C S H R L P A P H V R A M P S V P P A G
 S A Y Q P P V E M C H S F P P S Q G M A Q D A P I A Y Q R P M S E Q C L Q F P P Q
GCTTCAAGCAGGAGTTCATGACCCACAGTATGAGCAGGCGGGCAGCAGCTGCCAGTACCAGCAGCCGTGGTTCATCAAGCAGGAGCAGGTGGACTACATGTATGACTCAGGTGGGCAAAGAG 10332
 A S S R S S M T H S M S R R A A A A S T Q Q P W S S S R S R W T T C M T Q V G K E
 L Q A G V P . P T V . A G G Q Q L P V P S S R G H Q A G A G G L H V . L R W A K R
 G F K Q E F H D P Q Y E Q A G S S C Q Y P A A V V I K Q E Q V D Y M Y D S G G Q R
 AGCTTTTCTGACCGTGGCATTGGCATGGATGAGCCAAGTCCATGGGTGAACATCCATCACTGGTGGGTCTGCGCTTTCCAATTGACTTTTGCATCTTCAAGGAGCTGGGAAGTGGCTGTG 10455
 S S F . P W H W H G . A K S M G E H P S L V G S G L S N C T F A S S G A G K V A V
 A L S D R G I G M D E P S P W V N I H H W W V L A F P I V L L H L Q E L G R W L C
 E L F L T V A L A W M S Q V H G . T S I T G G F W P F Q L Y F C I F R S W E G G C
 TAATGGGAATGCAAGCAGTTCTGGCTGCTTGGGGCTCCAAGAGGAGGTGAGGGCATTCACTTTCTGTAAGAAGTTATAGGAAGAAATTTCTGATAGGTTAGGAACATAGATGTAGGAGCTCA 10578
 . W G M Q A V L A A W G L Q E E V R A F T F C K K L . E E I S D R L G T R C R S S
 N G E C K Q F W L L G A S K R R S G H S L S V R S Y R K K F L I G . E L D V G A Q
 V M G N A S S S G C L G P P R G G Q G I H F L . E V I G R N F . . V R N . M . E L
 GAGGTGAAGATAACCCACTCAGACTAGATTTAAGGCATGCTTCTGTGTTTTGCAAGAGAATAAGCAGCCTTTGGAGGAGGCTTGAATTCACTGATTAACCCAGAAGAGCTTGGCTTAT 10701
 E V K I T P L R L D L R H A S C C F A R E . A A F G G G F A I H . L T Q K S L A Y
 R . R . P H S D . I . G M L L V V L Q E N K Q P L E E A L Q F T D . P R R A W L I
 R G E D N P T Q T R F K A C F L L F C K R I S S L W R R L C N S L I N P E E L G L
 TCTTAAGGCTGTTTTGGCAGCCCTGGAGGTCAAACGTTCTTCTTGCCTCATGCTGAAGTTGAGTGCCTTTGGACAACAGCTGATTTTGACAGGATGAGAAGCCTGTTCTGACTTAGTGTCT 10824
 S . G C F G S P G G Q T F F L A S C . S . V P L D N S . F . Q D E K P V L T . C S
 L K A V L A A L E V K R S S L P H A E V E C L W T T A D F D R M R S L F . L S V L
 F L R L F W Q P W R S N V L P C L M L K L S A F G Q Q L I L T G . E A C S D L V
 CCCTGCCACTCCATTCTGGTGAAGCTGCTCCAGAAACTGCAGGGTTGTGCTCATCATTACATCCCAAGCTCTTTGGTGAAGTCTGAAGTGGGCTTCTATCCCCCTCCCCGCCACCTTTCCAG 10947
 P C P L H S G E A A P E T A G L C S S L H P K L F G E V . T G L L L F P S P P P F Q
 P A H S I L V K L L Q K L Q G C A H H Y I P S S L V K S E L G F Y S P P R H L S R
 S L P T P F W . S C S R N C R V V L I I T S Q A L W . S L N W A S I P L P A T F P
 AGCAGTGTGGGATGTGGGATCACATTATCAAAAACAAATCTATAAGCATGAGAAAGGAGCCCTTTGTGTTGTAGGTAATTA AAAACCTGGAAAAGACTNCTAAAATGGTCCATGCAA 11070
 S S V G M W D H I H Q K P N L . S M R K E P F V F V G N . K P G K D ? . K W S M Q
 A V W A C G I T F I K N Q I Y K A . E R S P L C L R . V I K N L E K T ? K N G P C N
 E Q C G H V G S H S S K T K S I K H E K G A L C V C R . L K T W K R L L K M V H A

— Exon 8

TGCCATACCTCATCTCAGCCAGGCTGGGTCAACTGCATTAATAAACACCACTGATTTTCATTTGCATTAATGGAACCAAACCGGCTTCTTGGGCTGCAGTCCCTAAAAGCCACTTGGGCTCT 11193
C H T S S Q P G W V N C I N K H H . F H L H . W N Q N R L L G L Q S L K A T W A S
A I P H L S Q A G S T A L I N T T D F I C I N G T K T G F L G C S P . K P L G P L
M P Y L I S A R L G Q L H . . T P L I S F A L M E P K P A S W A A V P K S H L G L
GCCAGCCTGGTGCATTGGGAGGTTGGACAAGAATCTGCATCTTAGGTATATTTTTATCTGTGCAGGTAATGCTGATCATTCTGGTTGGATTCCAAGGGCGAGTCCCTAAAGGCAGGAG 11316
A Q P G A F G R F G Q E S C I L R Y I F I W S G N A D H S G W I P R A S P . R Q E
P S L V H L G G L D K N L A F L G I F L S G Q V M L I I L V G F Q G R V P K G R S
C P A W C I W E V W T R I L H S . V Y F Y L V R . C . S F W L D S K G E S L K A G
CTTGGCTCTGCCAGGACTGGCAGAGAACTGGAATCAGGGCTGTCTGCTCATTCTTCTGCCCTAGAAGTGTGTAAGTTTTCTCCTATCTCAGTAGCAATGCATTTCCACTGCCATAGAA 11439
L G S A R G L A E K L E S G L V C F I F C P R S V V S F L L S Q . Q C I S T A I E
L A L P E D W Q R N W N Q G L S A S F S A L E V L . V F S Y L S S N A F P L P . N
A W L C Q R T G R E T G I R A C L L H F L P . K C C K F S P I S V A M H F H C H R
CAAGGAATGATGCAATGGAATGTACAGCCTAGGGAAGTGTAGCCCCAGGGCCAAGGAATGATAGGAGCTGTTGTCTTGTCTGCCCGGCTGTGATCCGGATAAAACATCTGTGCAAGGT 11562
Q G M M Q M E M Y S L G N C . P Q G Q G M I G A V V F A A R S V I R I K H L C K G
K E . C K W K C T A . G T V S P R A K E . . E L L S L L P G L . S G . N I C A K V
T R N D A N V Q P R E L L A P G P R N D R S C C L C C P V C D P D K T S V Q
GCAAAACAAATGACACAAGTAGATAACCTTCTTTTAGCTTCCCTGCAGCCGACAATGTGCACTGTGTTAAGGGGACCTTTTTCTCCCTTTTTTAGCTCATTTTTT 11685
A N K C N N G H K . I T F L L A S L Q P D N D V H C V K G T F F S S P F F S S F F
Q T N A T M D T S R . P S F . L P C S R T M M C T V L R G P F F P P P F L A H F L
C K Q M Q Q W T Q V D N L P F S F P A A G Q . C A L C . G D L F F L P L F . L I F
GTCTATCTCCTTGGTCATACGAGGCATCAGAAACAGGATCTGGGGGAATCTCAAGACTTCTGCTATTCACTTCTTCAGAAGGCATTTGGAGACCATAGTGGTCTACATGAAGGCTCCGA 11808
V Y L L G H T R H Q K Q D L G G I S R L L L F T S F R R H L E T I V V L H . R L R
S I S L V I R G I R N R I L G E S Q D F C Y S L P S E G I W R P . W S Y I E G S D
C L S P W S Y E A S E T G S W G N L K T S A I H F L Q K A F G D H S G P T L K A P
CACATTAACCTCATCTGGTTAGAGAGCTCTGTGCAGAGAAGCTCCAACACTGGGCGCTGAAGCTTGCATTGAAGGCATTAGAAAAGGAGAGGGGAAAAGGTGTTCCAGCTTCAAGGGGAGT 11931
H I N S S G . R A L C R E L Q H W A L K L R I E G I R N G G E G K G V P S F K G S
T L T H L V R E L C A E N S N T G R . S F A L K A L E M E E R E K V F P A S R G V
T H . L I W L E S S V Q R T P T L G A E A S H . R H . K W R R G K R C S Q L Q G E
TCCCTCTTGGGGGATGTGGAGGCTGGCAGGAAGCAGAGAGGGAGAGATGTGAGATCAGTGTGGATGGAATTATACCAAAAACCTTACCATTAACAGCAGAAAAGGAAGGAATCCAATAT 12054
S P L E G D V E A G R K Q R G R D V R S V W M E L Y Q K P S T I N S R K E G I Q Y
P L L R G M W R L A G S R E G E M . D Q C G W N Y T K N L L P L T A E R K E S N I
F P S . G G C G G W Q E A E R E R C E I S V D G I I P K T F Y H . Q Q K G R N P I
CTCAGAGCCAGTGTGGAACATCCACCTCAGGTCTCTGCAGCAGAGGTGTCTTAAGTGTCTGAAGTGCAGGGGTTGTGCTATACTATAAGATGATGCTATGGGGTAGTGCTATGGGATGACA 12177
L R A S V E H P P Q V S A A E V S L S C L K C R G C A I L . D D A M G . C Y G M T
S E P V W N I H L R S L Q Q R C L . A A . S A G V V L Y Y K M M L W G S A M G . H
S Q S Q C G T S T S G L C S R G V S K L L E V Q G L C Y T I R . C Y G V V L W D D

CTCTGTGCCCTCCTGGCCCCGTTCAAACCTAACCTATAAACTGACAACTGAGCCTGTTGCCCCAGACTTGATGTGAAGGCTGTTCTGCAGGAAGCCAGGTAGCACTGTGGAGATCAGCTG 12300
L C A L L L A P F K P . P Y K L T N . A C C P R L D V K A V L Q E A R . H C G D Q L
S V P S W P R S N P N P I N . Q T E P V A P D L M . R L F C R K P G S T V E I S W
T L C P P G P V Q T L T L . T D K L S L L P Q T . C E G C S A G S Q V A L W R S A
GTTCAATTCATGGTACCAAACGTGGTAAAAGTGAAGTTGTAGTGGAGCAGCTGGGCTGCAAGAAGGGATCTGTCTGTCTTCTGCCCTTCGTGGGTGCTGCTAAAACCTGTGCTTGTGCTGCCAT 12423
V Q F M V P N V V K V K L . W S S W A A R R D L S V F C P S W V L L K L C L C C H
F N S W Y Q T W . K . S C S G A A G L Q E G I C L S S A L R G C C . N C A C A A I
G S I H G T K R G E S E V V V E Q L G C K K G S V C L L P F V G A A K T V L V L P
ATCTGCTTCTGACCTCTGTGCCTTCCCTCCCTCCACGCTGACGCTATGACAAGCAGATGAGGCCCTTTGCTGACGATGTCTGTGTTGTTCCAGAAAAATTTGAAGGTTAGAGGAAGAATAA 12546
I C F . P S V P S F P S T L Q L . Q A D E A L C . R C L C C S R K I . R L E E E .
S A S D P L C L P S L P R C S Y D K Q M R P F A D D V C V V P E K F E G . R K N N
Y L L L T L C A F L P F H A A A M T S R . G P L L T M S V L F Q K N L K V R G R I
CTTCTTTAAATGGGGTTATCCAGCATCTCTGCTAGCAGCTTGGCCACTGCCTGCATTGCTCATAGTGCATTACTGTGATTGATGAGCGGCTCTTTGGCACCTGTTTCTGAGCATCTTTAG 12669
L L . M G L S S I S A S S L A T A C I A H S A F T V I H E R L F G T L F P E H L .
F F K W G Y P A S L L A A W P L P A L L I V H S L . F M S G S L A P C F L S I F S
T S L N G V I Q H L C . Q L G H C L H C S . C I H C D S . A A L W H P V S . A S L
CAGCAGTTGTAATGATCTTAATTTGGGAAGTGCATCAGAGCTTCTAGAAGAGCTGCTGGTGAATTTGCTGCTTTGGTTGTGGAAGTGAATATGCCGGTAAACTTTGGTGAGATG 12792
Q H V V M I V S . L G S A S E L L E E L L V N L L L W F V . T E I C R V N F G E M
S T L . . S Y L N W E V H Q S F . K S C W . I C C F G L C E L K Y A G . T L V R C
A A R C N D R I L I G K C I R A S R R A A G E F A A L V C V N . N M P G K L W . D
TGCTCTCTTGCAGCTGTGTTTTACCCTCAGAAGCACTTTGTCCCTCTCAGCAGGTGACATCAAGCAGGAGGTGCGAGGGTACCGGAGGGCCCTCCCTATCAGCGGAGGGATCTCTCCAGC 12915
C S L C S C V L P S E A L C P S Q Q V T S S R R S E G T G R A L P I S G G D L S S
A L F A A V F Y P Q K H F V P L S R . H Q A G G R R V P G G P S L S A E G I S P A
V L S L Q L C F T L R S T L S L S A G D I K Q E V G G Y R E G P P Y Q R R G S L Q
TCTGGCAGTTCCTTGTGGCTCTATTGGATGATCCCACTCCCACTTTCATTGCCTGGACTGGAAGAGGGATGGAGTTCAAGCTGATTGAGCCAGAAGAGGTGAAGCCTTCTGTTTCTCTG 13038
S G S S L W L Y W M I P P T P T S L P . G L E E G W S S S . L S Q K R . S L L V S L
L A V P C G S I G . S H Q L P L H C L D W K R D G V Q A D . A R R G E A F L F L C
L W Q F L V A L L D D P T N S H F I A W T G R G M E F K L I E P E E V K P S C F S
CAAGCTCAGGGGGAGCAGCTGCTGACCTCTGAGAGCCTCCCAAGTCCAACCTCAGCCACCCCCACCGTGCCCAACACATCCCTCAGTGTCCATTTTCACTGTTCTTAATGTCCTGC 13161
Q A Q G E S Q L L T S E S L P K S N L S H P P R A P P T T S L S A P F S L F L M S C
K L R A S P S C . P L R A S P S P T S A H P P R A H Q P H P S V L H F H C S . C P A
A S S G G A A A D L . E P P Q V Q P Q P T P T V P T N H I P Q C S I F T V L N V L
AGGCAATTGGAAGTCTGAGCCCGTGTATTGTTGAGCCTTCTTAATGCTGTAATGTTTCTATGTGCTGCGGTCCTTCTATCTGAAGATGATCAGCTGGATGGTGTCTTTGTGCT 13284
R Q L E L L S P C Y C L S L F L K C C K C F Y V L R S P S S E D D Q L D G V L C A
G N W N C . A R A I V . A S S L N A V N V S M C C G P L H L K M I S W M V F F V L
Q A I G T A E P V L L F E P L P . M L . M F L C A A V P F I . R . S A G W C S L C

— Exon 10 — Exon 11

CACAGCCACTCACACTCTGTGGCTTTATCTCTGCAGGTAGCCAGGCTGTGGGGCATCCAGAAGAACCGGCCAGCCATGAACACTGACAAGCTGAGCCGCTCTCTCGCTACTATTATGAGAA 13407
H S H S H T L W L Y P A G S Q A V G H P E E P A S H E L R Q A E P L S S L L L . E
T A T H T L C G F I L Q V A R L W G I Q K N R P A M N Y D K L S R S L R Y Y Y E K
S Q P L T H S V A L S C R . P G C G A S R R T G Q P . T T T S . A A L F A T I M R
AGGCATCATGCAGAAGGTGAGTTTGCAAAGCATCATACAATGGCTGGGTTAAAAAGGACCACAATGATCATCCAGTCCAACCCCTGCTGGTGGCAGCCCTGCACATAGCAGACCAGGCT 13530
R H H A E G E F A K H H T M A W V K K D H N D H P V P T P C W L A A L H I A D Q A
G I M Q K V S L Q S I I Q W P G L K R T T M I I Q F Q P P A G W Q P C T . Q T R L
K A S C R R . V C K A S Y N G L G . K G P Q . S S S S N P L L V G S P A H S R P G
GCCAGAGCCACATCCAGCCTGGCCTGAATGCCTGCAGGGATGGGGCATCCACAGCCTCCTGGGCAACTGTTCCAGTGTGTACCACCCTCTGGGTGAAAAAATTCCTCCTCATATCCAA 13653
A Q S H I Q P G L E C L Q G W G I H S L L G Q P V P V C H H P L G E K I P P H I Q
P R A T S P L A L N A C R D G A S T A S L G N L F Q C V T T L W V K K F L L I S N
C P E P H P A W P . M P A G M G H P Q P P W A T C S S V S P P S G . K N S S S Y P
CCTAAACCTCCCCTGTCTCAGTTTAAACACCTCCCTTTTGTCTATCACCATCCCACCTCATAAACAGAGCAGCTGGAGCTTTTGTGCCACTTGTGATGATCCCTTCTGGGAGAGCCAGGA 13776
P K P P L S Q F K T I P F C P I T I P P S . T E Q L E L L C H L . . S L L G E P G
L N L P C L S L K P F P F V L S P S H P H K Q S S W S F C A T C D D P F W E S Q E
T . T S P V S V . N H S L L S Y H H P T L I N R A A G A F V P L V M I P S G R A R
AGCCTTGCAGGTGCACCATGTCTCGGGCTCACTGATATACTCATCCCTTATCCAGCTCTGCAGCAGGTAATGAGGATGGTCAAAGGAGCTGGAGCCAGGAATGACATAAAGAAATGATG 13899
S L A G A P C P R A H . Y T H P L S T L S A S R . M R M V Q R S W S Q E . H K E M M
A L Q V H H V L G L T D I L I P Y P R L A A G K . G W C K G A G A R N D I K K . W
K P C R C T M S S G S L I Y S S L I H A L Q Q V N E D G A K E L E P G M T . R N D
GCTCCTTTGTATTTCTGCATATCAGTAGTAAGAAGAGTTTCAATTTAGGCAGCTGCAAAGGTGTACAATCAGGTGAGGGAGAGACCCACTGCATGTGGCTTGGCTCTGCATGGTGGGGA 14022
A P L C I F L H I S S K K S F N L G S C K G V Q S G E G E T H C M W L G S A W W G
L L F V F S C I S V V R R V S I . A A A K V Y N Q V R E R P T A C G L A L H G G D
G S S L Y F P A Y Q . . E E F Q F R Q L Q R C T I R . G R D P L H V A W L C M V G
TGTTCTCTCCAGATCCAGGACATTGGAGGGGATAACTGTGCGTGTGACACCGTCTGGGCAATGAAGCTGTGGCTGCATCCTCTGCTCTTTCTCCTACAGGTGGCGGGGAACGCTAC 14145
C S L P D P G H W R G . L C V . H R A G Q . S C R L H P L L S F S L Q V A G E R Y
V L S Q I Q D I G G D N C A C D T V L G N E A V G C I L C S L S P Y R W R G N A T
M F S P R S R T L E G I T V R V T P C W A M K L S A A S S A L F L P T G G G G T L
GTCTACAAGTTTGTGTGAGCCCGAGGCTCTGTTCTCTGGCCTTCCTGACAACCCAGCGCCGCGCCCTGAAGGCAGAGCTGGAGAGGCAGATCAGCGAGGAGGACACGGTGCCTCTGTCC 14268
V Y K F V C E P E A L F S L A F P D N Q R P A L K A E L E R Q I S E E D T V P L S
S T S L C V S P R L C S L W P S L T N S G R P . R Q S W R G R S A R R T R C L C P
R L Q V C V . A R G S V L S G L P . Q P A A G P E G R A G E A D Q R G G H G A S V
CACCTGGATGAGAGCACTGCTTACTGCCTGACCTTGGGAGCCTCCCCCGCAGCTCTATGGCAAAGCTACACCTACTAGTGTCTGGAGCTGTGTGCCACCCTGATCGTGATATAAGGA 14391
H L D E S T A Y L P D L G S L P P Q L Y G K G Y T Y . C C W S C V P P . S C I . G
T W M R A L L T C L T L G A S P R S S M A K A T P T S A A G A V C H P D R V Y K E
P P G . E H C L P A . P W E P P P A A L W Q R L H L L V L L E L C A T L I V Y I R

— Exon 12 — Exon 13 Putative Stop Codon Note: The end of the reported cDNA sequence is shown

GATGCTGGTTAGTAACCAAGCTGTATAGGTGTTTTCCAGGCCAGCCTGGGGCTGCACAGAGCTATGATACCTGCTGAGGGCAGGAGAAAGTTTCAGGTCAATGTGGCTGGATAACCAGC 14514
D V W L V T K L Y R C F P R P A W G C T E L . Y L L R A G E R F R S M C G W I T S
M S G . . P S C I G V F P G Q P G A A Q S Y D T C . G Q E K G S G Q C V A G . P A
R C L V S N Q A V . V F S Q A S L G L H R A M I P A E G R R K V Q V N V W L D N Q
AGGGGGTCAATGGGGCTGCTGGGTGCTGGTGGTCCCAACCCATATCTGGCAGGATGGGACCACTCCTCAGTGTGAAGTCATCACCTGAGCAGGGCAGCTGTGGCTTCCCAGCATCACAGGTG 14637
R G S M G L L G A G G A N P I S G R M G P L L S V K S S P E Q G S C G F P A S Q V
G G Q W G C W V L V V P T P Y L A G W D H S S V . S H H L S R A A V A S Q H H R W
Q G V N G A A G C W W C Q P H I W Q D G T T P Q C E V I T . A G Q L W L P S I T G
GTCGCTGCCTCCAGCAGGAGCCACCTACAGCATTCTCCCCCTTGGCAGATGTGATGCCATCCTGCTTGGCTTCTTCTCACAGCTGGGGATGCTACAGCTGCTGCTGGCTGCCTTCTCCA 14760
V A A S S R S P P T A F L P P W Q M . C H P A W L L L T A G D A T A A A G C L P P
S L P P A G A H L Q H S S P L G R C D A I L L G F F S Q L G M L Q L L L A A F L Q
G R C L Q Q E P T Y S I P P P L A D V M P S C L A S S H S W G C Y S C C W L P S S
GTGCTGACCCCTGGGTATATTGCCCTCAGCCGGGGCAGCCAGACCAGTGCAGATTCACCTCCTCAGGGCCAACCTTCTTCTGCCAGGAGAGC 14857
V L T P W V Y C P Q P G Q P R P V Q I H L L R A N L L F C P G E
C . P P G Y I A L S R G S P D Q C R F T S S G P T F S S A Q E S
S A D P L G I L P S A G A A Q T S A D S P P Q G Q P S L L P R R A

Appendix B The Pufferfish PEA3 5' RACE Product

The nucleotide sequence as well as the deduced amino acid sequence is shown. The exon borders have been indicated by vertical line, and the in-frame start codons in exon 2 are boxed in green.

CAGGTCAAACGGTTAACGGATGGATTATAAGATGGATGGATATTTGGACCAGCAAGTGCCTTACACTTTAGCAAATAAATCGCATGGAAATGGGCCCTAAACAGACTGTTAATGGCAGC 120
R S N G . R M D Y K M D G Y L D Q Q V P Y T L A N K S H G N G P L N R L L M A A
Exon 2 Exon 3

CAAGAGGAAATACATGGAAACAGAGCTGCCACCTCAGGAATCCGAAGATCTTTTTTCAGGATTTAAGCCAACCTCCAGGAGACGTGGCTCACAGAAAGCTCAGGTTCTGACAGCGACGAGCA 240
K R K Y M E T E L P P Q E S E D L F Q D L S Q L Q E T W L T E A Q V P D S D E Q
Exon 3 Exon 4 Exon 4 Exon 5

GTTTGTCCCGACTTCCAT 259
F V P D F H

Appendix C Pufferfish PEA3 Genomic Sequence at the 5' End of the Gene

The genomic sequence of the 5' end of pufferfish PEA3, including 1 kb of sequence upstream of the putative start codon in exon 2, is depicted. Exons 2, 3 and 4 have been underlined. The two in-frame start codons in exon 2 are indicated.

GGTACTTTGGATCGGACCGCTCTGGAGCTGCAGCCTCTGGAGTTGCAGCCTCTGGAGCTGCAGCCTCTGGAGCTTTAGATCTCGCGTTTCCGTTGGCGCAGACCTAAATAAAAAACCTCAG 124
G T L D R T A S G A A A S G V A A S G A A A S G A L D P A V S V W R R P K . K T S
V L W I G P P L E L Q P L E L Q P L E L Q P L E L . I L R F P C G A D L N K K P Q
G Y F G S D R L W S C S L W S C S L W S C S L W S F R S C G F R V A Q T . I K N L R
GAATTCAGAATCTTCAGGATTGCGTCAGTTTGGGAGCTGATATAAAGGGCTGCTTTTTTTCTTCTATTTTGTCTGTTTTTCATTTTAAAAACAACACTGACGCGAGTCAAAGTTAAAAACA 248
G I Q E F F R I A S V W E L I . R A A F F F F Y F V C F S F . N N T D A S Q S . N K
E F K N S S G L R Q F G S . Y K G L L F F S S I L S V F H F K T T L T R V K V K T
N S R I L Q D C V S L G A D I K G C F F F L L F C L F F I L K Q H . R E S K L K Q
ACAAAAAAGATAGATAGAGAGAAAAACATGTCGGGAGTTTTCTGTCAGGGCAGGTTTTATGAATGGAGCCCTTTCTTCTCTCTCTCTTTTGTGAATGAAAAACAACGGAAAA 372
Q K K K I D R E K N M S G V F V P G Q V F M N G A L F F L S L S F C E . K T T G K
N K K K R . I E R K T C R E F S C Q G R F L . M E P F S F S L S L F V N E K Q L E K
T K K K D R . R E K H V G S F R A R A G F Y E W S P F L S L S L F L . M K N N W K
GCCACAATAAAACCCAGCAGCTGATTGGCTGCCGCGCAACCTCGGCCCGCCACCGAGTTATCGTCTCATTCCAGCACTCCGATTGGCCCGCGTGGGAGAGGAGGGGGGGTTCATTCAT 496
A T I K P S S . L A A A Q P R P R P P T L S S S F Q H F R L A R V G R G G G V H S
P Q . N P A A D W L P R N L G P A H R R Y R P H S S T S D W P A W G E E G G F I H
S H N K T Q Q L I G C R A T S A P P T D V I V L I P A L P I G P R G E R R G G S F I
AAGTGAGCGTATTCAAGGGGACAAAGGGGAGAAATTCAGGCTCTTGAAGAGGAAATCCACCAATCAGGAGGCGAAGCAACTTTATCCCCGAATCTGTCCGGCAGGCGCGTGCAGGTGCAAAAC 620
. V S V F K G T K G R N S S S . K R G S T N Q E A K N L Y P R I C P A G A V T V Q T
K . A Y S R G T Q R G E I R A L E R K N P P I R R R S N F I P E S V R Q A P . R C K
S E R I Q G D K G E K F E L L K E E I H Q S G G E A T L S P N L S G R R R D G A N
TCTAGATTATACGTCAGAAATGCGTAAAAAGCTGTTCAACCTGTTGACTGCGGGGAGATAAACGCGCTGAAAGCCGAGCCGTGCACCCCGCCCTCCCGACCTCAGCACGCGCCGT 744
L D Y I R P E M R K K A V Q P V D C G E I N A L K A E P C T P R P L P D L S T A R
L . I I Y V Q K C V K K L F N L L T A G R . T R . K P S R A P P A P S P T S A R P V
S R L Y T S R N A . K S C S T C . L R G D K R A E S R A V H P P P P P R P Q H G P
CTGACCCGTTTACGCACAAAGTCTCCGAAAGTCCGGCGCGGAGGTGCGAAACTTCAGCGCCGTCGTCACACATGCGCGACCCACAAACAAAAAAGGAGGCTGTGTAGGGGTG 868
L T R L R T K C S E S P A R R C E T S . A P V R H T C A H P Q T K K K K A G L C R G
. P V Y A Q S A P K V R R G G A K L Q R P S V T H A R T H K Q K K K R Q V C V G V
S D P F T H K V L R K S G A E V R N F S A R P S H M R A P T N K K K K G R S V . G W
GGGGCAGCGCCCGTGGTAATCACTTCCCTCGGTGTTTTTTTTGTGAATGGAGCTCATTATGCATGCAGCCTCCACGCGAGTCCATTGACGCGAGTCCCTCCTCTCATCTGAGTCCC 992
G G Q R P A G . S L P L G V F F C E W S S L C M Q P P R S S I D A Q L P P L I . V P
G G S G P L G N H F P S V F F F V N G A H Y A C S L H A A P L T R S S L L S S E S
G A A A R W V I T S P R C F F L . M E L I M H A A S T Q L H . R A A P S S H L S P
TGGAGCTGCGCGGACGCTCAGCCCCGATGACCACAAACGGTTAACGGATGGATTATAAGATGGATGGATATTTGGACCAGCAAGTGCCTTACACTTTAGCAAATGTACGTATGGATGCAGC 1116
A S C A D Q L S P D D H K R L T D G L . D G W I F G P A S A L H F S K C T Y G C S
L R A A R T S S A P M T T N G . R M D Y K M D G Y L D Q Q V P Y T L A N V R M D A A
C E L R G P A Q P R . P Q T V N G W I I R W M D I W T S K C L T L . Q M Y V W M Q

— Exon 2 (including up to the end of the 5' RACE product) In-frame Start Codons

TCGACCGTCCGGCTCGGCACGGCTTTATTTGTTTTAATCGTCGCGTCCGCGTCCGAGCGAGTTTTGCTCCGTGGCAGCAGCCCTGTGTGTGCGCGTGTGTGTGCGCGCGGAGAGCGAGTG 1240
S T V R L G T A L F V L I V A S R V R A S F A P W Q Q P C V C A R V C A R A R A S
R P S G S A R L Y L F . S S R R A S E R V L L R G S S P V C V R V C V R A R E R V
L D R P A R H G F I C F N R R V A R P S E F C S V A A A L C V C A C V C A R E S E C
TGC GTGTGTGTTTGTGTGCGACGCCGAACCTCGTCTAAACTAAGTAAAATGTTGCGCTAACGTCAATCACGCGCCGGGTTTGGTGATGTTTGGGGAGCCGAGCGGCCACTTTCCTCGCAG 1364
V R V C V C V R R R T R P K T K . K C C A N V N H A P G L V M F G E P S G H F P R R
C V C V F V C D A E L V L K L S E N V A L T S I T R R V W . C L G S R A A T F L A
A C V C L C A T P N S S . N . V K M L R . R Q S R A G F G D V W G A E R P L S S Q
GCGCGCTGTGTTTTGCTTTTTTCCCGCCGGAACGCGAAACGTGTCGCTTCTGAGCAGTTTGATGGTCACGCACGGCCGAGCCTTCCGCAGAAATGACAACAGCTGCCATTTTCGCTGAT 1488
R A S V L L F S P P R T R N V S L S E Q F D G H A R P Q P S A E M T T A A I F A D
G A R L F C F F P R R E R E T C R F L S S L M V T H G R S L P Q K . Q Q L P F S L I
A R V C F A F F P A A N A K R V A F . A V . W S R T A A A F R R N D N S C H F R .
TGATGCCCCCTTCTCACTTCTGCTCTCCGGGTTGTGCGAAGTGCATGGAATGGGCCCTAAACAGACTGTTAATGGCAGCAAGAGGAAATACATGGAACAGAGCTGCCACCTCAGG 1612
. S A P F S L L L S G F V Q K S H G N G P L N R L L M A A K R K Y M E T E L P P Q
D L P L S H F C S P G L C R N R M E M G P . T D C . W Q P R G N T W K Q S C H L R
L I C P F L T S A L R V C A E I A W K W A P K Q T V N G S Q E E I H G N R A A T S G
AATCCGAAGTAAACTCAGCTCAAATCTCAATACTTGACCTCGGCTCAGGTTATGATGAACCTCGGGACGTGTCTGAGTCGGACCCAGCTTCTCTAAACTGTCTCCCCACTCCCGATCT 1736
E S E G K T Q L K S N Q Y L T S A Q V M M N L G T C L S R T Q L L . T L S P H S P D L
N P K V L S S N L N T . P R L R L . . T S G R V . V G P S F S K L C L P T P Q I
I R R . N S A Q I S I L D L G S G Y D E P R D V S E S D P A S L N S V S P L P R S
TTTTCAGGATTTAAGCCAACCTCAGGAGACGTGGCTCACAGAAGGTGAGATTTCTCTTTGTTCTCTCTGCTCGCTGCTTCTGTGGACGATCTCTCATCGTGGTGATGGACTTAATCAGCC 1860
F Q D L S Q L Q E T W L T E G E I S S L F L F C S L L L W T I S S S W . W T . S A
F F R I . A N S R R R G S Q K V R F P L C F S S A R C F C G R S P H R G D G L N Q P
F S G F K P T P G D V A H R R . D F L F V S L L L A A S V D D L L I V V M D L I S
GCAGCTTCTGCTGGGCTTTTTTCCCTCCCAATGGGCGAGAAGGTGACCTCGTGTGTTGCTTAAACTTCTGACGGGACACATATGCTCATAAACCCCTGGATGGTCAGAAAGAGGCCG 1984
A A F L L G F F P S P M G E K V T S C V R S K L S . R D T Y A H K P L D G Q K E A
Q L S C W A F F P P Q W A R R . P R V F V L N F P D G T H M L I N P W M V R K R P
R S F P A G L F S L P N G R E G D L V C S F . T F L T G H I C S . T P G W S E R G R
GAGGGGAGAACACAGACGGGACAGTATTAGCTTGAGCATGTCCTATTTTGGCCATGTTGGCATGTGGCGTGCCTGCCAGCCAGCCAGCAGCCAGCCAGGGCAAAACAAACAGGCTGTTGGGCTCG 2108
G G E N T D G T V L A W S M S Y F G H V G M W R A C Q P A S S S Q G K Q T G C W A R
E G R T Q T G Q Y . L G A C P I L A M L A C G V R A S Q P A A A R A N K Q A V G L
R G E H R R D S I S L E H V L F W P C W H V A C V P A S Q Q Q P G Q T N R L L G S
TGTCACTCAGCAGCACAGAGATGGTCCAGGAGGTGGGAAGCCAGACTGGTGTATGGGGGGGGGGTGAACCGGAGGTGCGATTGTGTTGCTCGGCTGCTTTGATGCGCGTTTGTATGTC 2232
V T H A A Q R W S R R W E A R L V L W G G G . T G R C R L C C L G C F D A R L M C
V S L T Q H R D G P G G G G K P D W C Y G G G G E P G G V D C V A S A A L M R V . C A
C H S R S T E M V Q E V G S Q T G V M G G G V N R E V S I V L P R L L . C A F D V

— Exon 3 — Exon 4

AGGTTTAAAGGGGAGCGGGCTTCATTCTTGACATTTAATGTGAACCACATCGACATGATAAATCCTGTTTAAACCG 2310

R F K G E R G F I L D I . C E P H R H D K S C L N P
G L K G S A A S F L T F N V N H I D M I N P V . T R
Q V . R G A R L H S . H L M . T T S T . . I L F K P

Appendix D Pufferfish PEA3 Genomic Sequence Within the Gene

The genomic sequence of a large fragment of genomic pufferfish PEA3 containing exon 5 has been completed. The 5' border of exon 5 was contained within the 5' RACE nucleotide sequence and has been indicated within the sequence by beginning to box the deduced amino acid sequence at the valine amino acid that is coded by the first codon of the exon. The remaining amino acids carboxy-terminal to this valine residue are boxed until an in-frame stop codon is reached in the nucleotide sequence.

CCC GGGG CCGAAGAGGCAGAAAAGTTAAGAGGGGCTAATTCATTACGGAATAGTGGCCACAGGATGCGCCGCTGTCCAAGAAATTTGATGCCATTTATCCGAACATGTTCCAAAACGCCGTC 124
P G G R R G R K V K R G . F H Y G I V G H R M R R C P R N . C H L S E H V P K R R
P G A E E A E K L R G A N S I T E . W A T G C A A V Q E I D A I Y P N M F Q N A V
A R G P K R Q K S . E G L I P L R N S G P Q D A P L V S K K L M P F I R T C S K T P S
ATCCGGTTTCTCTGCGCAGGAAGCGCGCGCTGACGACAAAGCAGATTTGAATTTCTTGCTTCCCGTTTCTGCTCCTCTGGACAGTTTTGCTTTCTGTCTTGTCTCACCGGAGACCGGA 248
H P V S S A Q E A A A . R Q S R F E F L A S P F L V L L W T V L L S V L S H R R P E
I R F P L R R K R R R D D K A D L N F L L P R F S S S S G Q F C F L S C L T G D R
S G F L C A G S G G V T T K Q I . I S C F P V S R P P L D S F A F C L V S P E T G
AACGGAGGAAGCAACCGGGCGGACCAAGGCCGAGCCGTGCTCCTCCTCCCGGATCCTTACGGGGGGGAAAATGTAATTTGCAGAATTCAAAACGCCAAATTTATGTGAAATTAGAAGGT 372
T E E A T G R T Q G R S R V P P P G S F T G G E N V I C R I P N A K L C G N . K V
K R R K Q P G G P K A A A V S L L P D P S R G K M . F A E F Q T P N Y V E I R R F
N G G S N R A D P R P Q P C P S S R I L H G G G K C N L Q N S K R Q I M W K L E G
CTTTGATCTCCGACGTGATGGACAGAAGTGGCTGAAGGACGCGGGTTGGTTCAACCTAAATGTTGACGTCCACTCCAGGAGGACGTCCACTTCTGGAGGACGTCTCCAGGAGGTGGACG 496
L . S P T . W D R S G L K D A G L V Q P K C . R P L Q E D V H F L E D V L Q E V D
F D L R R D G T E V G . R T R G W F L L N V D V H S R R T S T S W R T S S R R W T
S L I S D V M G K W A E G R G V G S T . M L T S T P R G G R P L P G G R P P G G R R
TCCACTTCTGTCTGTTCTCTGAAAGCTCCGAGCTTCTTTATCCAAGAGTCTGCGGACCAACGTGTTCCCTTAATCTTCGATGAAGTCCAGACTGCATCCGGATCGTCTGCTGGTCCA 620
V H F L S R S L K A P S F F I Q E V L A D Q R V P L N L R . S P D C I R I V L L V Q
S T S C L V L . K L R A S L S K R F W R T N V F P L I F D E V Q T A S G S S C W S
P L P V S F S E S S E L L Y P R G S G G P T C S P . S S M K S R L H P D R P A G P
GAGACCAGCGGAGAGAAATCCCTCACTTGTCCGTTTGTAGTCTGCGGTGGGAATAAGAAACCTGAAACATGGAAGCAACTAATGGTGGTGAATGGAGACCAACTGGGATGATGACTGG 744
R P A E R E S L T C S V C S P A V G I R N L E T W K Q L T G G E W R P T G M M Y W
R D Q R R E N P S L V P F V L R W E . E T W K H G S N . L V V N G D Q L G . C T G
E T S G E R I P H L F R L . S C G G N K K P G N M E A T N W W . M E T N W D D V L
GGTGCCTCTTCTCTGGTGTCTTAGTTGCCCTTCCAGGTTGGAGGGTGGAGTTGCCATTATTGCTCATTGATTAATGTCGGGAAGGTGAGCCGGAGGTCGTGTGGGAGAACTTC 868
G A S F L W C S . L P P S R L E G W S . C H Y C S L I N V P G R S G R R S C V E K L
V P L S S G V L S C P L P G W R G G V A I I A H . L M S R E G Q A G G R V W R N F
G C L F P L V F L V A P F Q V G G V E L P L L L I D . C P G K V R P E V V C G E T S
CCTCTAATCCCGTTTCTTCTCTGCTCCAGCTCAGGTTCTGACAGCGACGAGCAGTGTGTTCCCGACTTCCATCCGAGAAGTGTGAGTAACTCCTTCTGTTTTTCCCGGCAACCGC 992
P S N S R R F F S A S S S G S . Q R R A V C S R L P F R E L . V T P S F F S R Q T R
P L I P V S S S L L P A Q V P D S D E Q F V P D F H S E N C E . L L R F F P G K R
L . F P F L L L C F Q L R F L T A T S S L F P T S I P R T V S N S F V F F P A N A
TCCGAAGCAGCGTCCAGCGGGTGTAGCTCCGCTTCTAGGGTCTATTTTTTTGTGGGGGGTGGTGGCAGACGCGAGCGGCTAATGAGCCGACATCGGGGAGCGCATGTGGAAGCCGACATC 1116
P K Q R P G G L A P L L G S I F F V G G W W Q T R A A . . A G H R G A H V E A D I
V R S S V Q A G . L R F . G L F F L W G G G G R R E R P N E P D I G E R M W K P T S
S E A A S R R V S S A S R V Y F F C G G V V A D A S G L M S R T S G S A C G S R H

Exon 5

CCGCGCTGCCCTCAGCCATTTGCATGCCTTTGAAGGGCTGAATGGGAACGACCGAGAAACCAGTAAAGAGACGGGCGAGGGGGCGCAGGTACAAAGATGGACCTTTGTGCCGCTCGGAAGACG 1240
P R C P Q P F A C L . R A E W E R R R K P V K R R A R G R R Y K D G P L C R S E D
R A A L S H L H A F E G L N G N D A G E T Q . R D G G G G A G T K M D L C A A R K T
P A A L P S A I C M P L K G . M G T T Q E N S K E T G R E G A Q V Q R W T F V P L G R
T G A A T G G A A G C A A A T T C T G G G A A G C C A G C T G A G C G G T C T T T T A G T G T T C C T T T C T G T T C C C A A C A A T G G C G G T T C A G G A T C A G G C C G G A G A C A A C C G T C C C C T T 1364
V K W E S K F L G S A A E R L F . W F P F L L V F P N N G G S G S G P G G Q P L P F
. N G K A N S W E A Q L S G S F S G F P F C L C S Q T M A V Q D Q G R E D N R S P
E M G K Q I P G K R S . A A L L V V S L S A C V P K Q W R F R I R A G R T T A P L
C C C A G C G G C C C G T C T C A G A A C A C C T C G G A A T A T C G C G G T T C A A C C A A A A C T G A C A G C T C T C A G T C A A A G C C A C T T C T G T T G C G A C T G C T G G C C C C G C C C T C G G T G A C A C A C G G C G 1488
P S G P V S Q N T S G I I A A S N Q N . Q L S V K A T S C C D C W P R P R . H T A
S P A A P S L R T P R E . S R L Q T D S S Q S K L P V A T A G P A L G D T R R
P Q R P R L S E H L G N N R G F K P K L T A L S Q S H F L L R L L A P P S V T H G
C C G T C G C T C T G T T G G C C A A C A A T G A G G A G G A A G T G C G A A T G A A C C C C A G G G C C G C G G G G C C A T G T C T A C T T A A C C C C T G A C C T G T C A G C C A A T C A G C T G T C A G G C C C C A C T A A C 1612
P S L C V G Q Q . G G G S A N E P Q G P A G A M S T . P P D L S A N Q L S G L H .
R R S V L A N N E E E E V R M N P R G R R G P C L L N P L T C Q P I S C Q A S T N
A V A L C W P T M R R R K C E . T P G A G G G H V Y L T P . P V S Q S A V R P P L T
G G T G A C C G G A C C T G A T A T G C T G G T C A C G G T G T G G G G G G G T C C G G T T G A C G G C A G C T T G T T A T C C G C A G C A G A C G C T G C G G C T T G A C G G A C G G C C G C T G A T G A G A C C C G C A 1736
R L T G A . Y A G S R V W G G P G . R A A C L S A A E T L R R . R T A A D E T P Q
G . P E P D M L G H G C G G G V R V D G Q L A Y P Q Q R R C G V D G R P L M R R R
A D R S L I C W V T G V G G G S G L T G S L L I R S R D A A A L T D G R . . D A A
G C G T C C A G C C C A G A A T C C G G A G C T G A C A A G T G A C A G A A A C A C G C G G T T C G T G T T A C A C A G C C C T C G C G C T C G T A A A A G T G C C G G G A T T A A A G G C T G T T C A G C G C T T C G T G T A G C A C 1860
R P A P E S G A D K . R E T R A F V F T R A L A L V K V P G F K G C S A R S C . H
S V Q R Q N P E L T S D E K H A R S C L H E P S R S . K C R D L K A V Q R V R V S T
A S S A R I R S . Q V T R N T R V R V Y T S P R A R K S A G I . R L F S A F V L A
A A T T A G C T T C C T T T G G T T G G C G G A A T G A C G T A G C G C G G C C C T T A A C G G C A G C C T G C G A A A G G T T A C A A C T G G T A A A A G T G T T C C C G G A G A C T G C A A A C A G C C C C A G C A G C T C T T T T 1984
N . L P L V W R E . R . R G A F N G S L R K G Y K L V K V F P G D C K Q P Q Q L F
I S F L W F G G N D A S A A P L T A A C E K V T N W . K C F P E T A N S P S S S F
Q L A S F G L A G M T L A R R L . R Q P A K R L Q T G K S V S R R L Q T A P A A L L
G T G C T C G G T C T G T T C A G T C T G T G T G T T G G T G T C A A A C A G G T G A C T C A C C T G T T T G G A A A C T C C G G C A A A A A A A G A G A C A G C A G T G G C A G A G C T G C G A T C C C A T G T G A G C A G G A T A T 2108
C A R S L C L V G V K T G D S P V L E T P A I K Q R A R G R A A I P C E Q D I
V L G L C S V C V W L V S K Q V T H L F W K L R Q . N K E H V A E L R S H V S R I
C S V C V Q S V S G W C Q N R . L T C F G N S G N K T K S T W Q S C D P M . A G Y
A A T T G A G G C T T G A A G A G A A G A A G A A G A G T A G A A G T C A A A T A A A C G T T G G C G T T A C C G A G T C A G G C T G C T A C T G C G C G C T G A G C G A C G G G A G G A G G C G C G G G A A A C A C G C G G T 2232
I E A . R R E E E E E V E R S N K T L A F T E S G C Y L R A . A T G G R R G K T A G
. L R L E E E K K K K . K G Q I K R W R L P S Q A A T C A R E R R R E G G A G K Q R V
N . G L K K R R R R S R K V K . N V G V Y R V R L L P A R V S D G R E A R E N S G

GGAAATGTTTGATTGGGAAGGGTTCGCGAACACCTGCACATCTGGTAAATCGATCTCATCGGCTGGTTTCTCGCTCAGCTCGCGCCCGACGCTGCTGCACAGGGTGTGTGTGTGTGTATGC 2356
G M F D W E G L R N T C T S G K S I S S A G F S L S S R P T L L H R V C V C V C M
E C L I G K G C A T P A H L V N R S H R L V S R S A R A R R C C T G C V C V C V C
W N V . L G R V A Q H L H I W . I D L I G W F L A Q L A P D A A A Q G V C V C V Y A
GTGTGTGGCGCGTGTGTGAGCCGATGATGATCGATGTCGTGGTGCCTTATTTTCATTTTCGCGCAGCCTTAAAGCAGCTCCAAACTCATCTTTATGAGCTCCTGTGATCCTTTGTGAGGGG 2480
R V C A R V . A D D D R C R G A P V F H F A A A F K Q L Q T H L Y E L L . S F V R G
V C A R V C E P M M I D V V V P P Y F I S P Q P L S S S K L I F M S S C D P L . G
C V R A C V S R . . S M S W C P R I S F R R S L . A A P N S S L . A P V I L C E G
CGGACCGCGCCGAGGTTCTTGCTGGGTCCCTTTGACCCGAAGGACCCCGAGGGTCTCTTCAGGGCCAAATGTTGCGGGTGTGCTGTCGCGACTGACTCTTTGTCGTCGCCCGCTGA 2604
G P G P R F L L G P L . P E G P P R V S F R A K C S V G C L L S D . L F C P S A A .
A D P A R G S C W V P F D P K D P E G S L S G P N V R G V C C P T D S F V R P P P D
R T R P E V L A G S P L T R R T P E G L F Q G Q M F G V F A V R L T L L S V R P L
CGGGAAGAAACACGCGGAGGTGCTACTCGAGTTTATGGTGCCTGAGTTGACGAGAGCGCTGTTACTGTGCGCGCTGCGCGCGCTGCTCTGGAGAGCGCTCCGCACAGGCTGCAGC 2728
R E E T R G G V T R V Y G A R S . R E R V V T V R A C A R A L P L E S A P H R L Q
G K K H A E V S L E F M V R V V D E S A S L L C A R A R A R C L W R A L R T G C S
T G R N T R R C H S S L W C A . L T R A R R Y C A R V R A R A A S G E R S A Q A A A
TTCTTTGTCCAACTATGTTGTTACACTTGGAGCAGCCCGTGGAGACGTGACCCATTATAAATTTTCTTNCCTGTTGCGCAAGAGGAAGCGTGTTCCTCCCCCCCCCCCCAGTCTCCG 2852
L L C P T M L F T L G A A R R R R D P F I N F L ? P C A Q E E A C F P P P P P V L R
F F V Q L C C L H L E Q P V G D V T H S . I F F ? R V R K R K R V S P P P P Q S S
S L S N Y V V Y T W S S P S E T . P I H K F S S P V C A R G S V F P P P P P S P P
CTGCTTCTCCCTCTCGCTCTTATTATAAAAAAAAAAGCCCTCCAGTACTAGAGGCTCCCTCGTCTCCAAGTGGACACACCAGGCCCTCCCTCCCTCTGTTTCTGCACAACTAAG 2976
C F S L L R S S F I K K K P S S T R G L P L V S K W T H Q A P P S L P V S A Q T K
A A S P S F A L H S . K K S P P V L E A S P S S P S G H T R P L P P F L F L H K L S
L L L P P S L F I H K K K A L Q Y . R P P P R L Q V G D T P G P S L P S C F C T N .
TAGAATTTAAAAATCGGACCCACCGCTTCCCTCCCTTTCTCTCTCTTCCACACACACACACACACACTCACGCACACACTCTCTCTCTCTCT 3100
. N L K I W T P T A S L P F P F S L S L S T H T H T H T H T H T H A H T L S L S L
R I . K S G P P P L P S P F P F L S L F P H T H T H T H T H T L T H T H S L S L S
V E F K N L D P H R F P P L S L F S L S F H T H T H T H T H S R T H T L S L S L
CTCTGAGTGTTTAAGATGGTGAACATAATCCAGAGCTGTGTGGGACGATTGTTGACTCGCTTGCAGTAATTATAGGAAGTTGTGGCTTAGAGCGGCTTACGGAGCACAACTGTAGC 3224
S L S V L R W . T N P E L C G G T I V . L R L A V I I . E V V A L E R R Y G A Q T V A
L . V F . D G E L I Q S C V G R L F D C V L Q . L Y R K L W L . S G V T E H K L .
S E C F K M V N . S R A V W D D C L T A S C S N Y I G S C G S R A A L R S T N C S
CGGAGAGATTCCTCTCGCTGGGATGGGATGGGATTCATAGGTGGGATTCATAGGTGGGATTTGCCCTTTTCTGACAGGGACAGCAACAAAACGTGATTTCTCGTCTGTAACACGCC 3348
G E I P S C V G M G G I P . V G F H R W D L P P F L T G T R T K T . F L V C K H A
P E R F P A L G W V G F H R W D S I G G I C A P L F . Q G H E Q K R D F S S V N T P
R R D S L L R W D G W D S I G G I P . V G F A P F S D R D T N K N V I S R L . N T R

CTGGTTTTCTTGACTTTTGTGTTGGAAATGCCAACAGATCTGTTCTGGCGTGTTCCTTTTGTTCGGTATTCTTCTCATTAGGAGGGTCACGGGTCACGAGCGTGACGCCGCG 3472
L V F L A L L L L E M P T D L F W R V S F C F S G I L S S F R R V T G H E R A A A
W F F L H F C C W K C Q Q I C S G V F P F V F P V F F P H L G G S R V T S V Q P P
P G F S C T F V V G N A N R S V L A C F L L F F R Y S F L I . E G H G S R A C S R R
TGACCCGGTGTCCAAAAAAGACCCGGTGTCCAAAAAAGAGGACCCGAAAAAGGGCGAGGAGGAGAGCAGACGCGAAATGGGGAAGAAAGTGCAGCTGAATCCTCCAGCAGAGCGCCTG 3596
V T R C P K K D P V S K K R G P E K G R G G E Q T Q R N G E E S A A E S S Q Q S A C
. P G V Q K K T R C P K K E D P K K G E E E S R R S E M G K K V Q L N P P S R A P
D P V S K K R P G V Q K K R T R K R A R R R A D A A K W G R K C S . I L P A E R L
TGTTCTGCTTAGTCATGAACCCGGCGCTTCAGCCTCCGTTGTTGTTCTGGATCAGAAGGGCCCTGTTTGGGCCCGATTGGAGCCCCCCCCCTCCCTCCTTGAGAGACCAGACGTGCAGCTCT 3720
V L L S H E P A A F S L R C C S G S E G P C L G P I G A P P L P S L R D Q T C S S
V F C L V M N P R P S A S V V V L D Q K G P V W A R L E P P P S P P . E T R R A A L
C S A . S . T R G L Q P P L L F W I R R A L F G P D W S P P P P L L E R P D V Q L
GAGTTCCTCCGGTTTTTCCAAAATACAGCTGCAACGCAACACTTCCAAAAAACCCAAAACGCAGCCAGAACTTACAGCGTCAACTGTGTTTTTGTCTGCTGGCCAGTAAATGGCCGTTTTAG 3844
E F S G F S K I Q L Q R N T S K K T Q N A A R T S D V N C V F V L L A S K M A V L
. S S P V F P K Y S C N A T L P K K P K T Q P E L Q T S T V F L S C W P V K W P F .
. V L R F F Q N T A A T Q H F Q K N P K R S Q N F R R Q L C F C P A G Q . N G R F S
CAGCAAACATGTGATTTCTTTCGCGCGCGCGCACACAGCGAGCGCGCGCTCTTCTGCAGTTTTATGAATGGGTGTGTTGTGAAAGAGCGGCTGAGCGGCACATTGAATTACAGGGAGT 3968
A A N M . F L S R A R A H T Q A R A L F L Q F Y E W V C C E R A A E R H I E L Q G V
Q Q T C D F F R A R A H T R R R A R S F C S F M N G C V V K E R L S G T L N Y R E
S K H V I S F A R A R T H A G A R A L S A V L . M G V L . K S G . A A H . I T G S
GAACCTTATTCTGACTTTTGGGCCACATCCCCCCCCCAATCACACACCACCGCTCGCGTCTCTCCGGTTAAAACGGTAAAAGTCTCCCGG 4065
N L H S D F W A T S P P P P I T H H R S R L S G . N G . K S P G
. T F I L T F G P H P P P P Q S H T T A R V S P V K T V K S L P G
E P S F . L L G H I P P P P N H T P P L A S L R L K R L K V S R

Appendix E Pufferfish PEA3 Genomic Sequence at the 3' End of the Gene

The genomic sequence of the 3' end of pufferfish PEA3, including a large amount of sequence encompassing the neighboring gene is depicted. The deduced amino acid sequences that bear homology with other PEA3 protein sequence and that corresponding to open reading frames in the genomic sequence are boxed. In this fashion the positions of exons 12 and 13 have been indicated. The putative polyadenylation signal is boxed in red and the open reading frame containing the final exon of the pufferfish HRH1 homologue is underlined.

CACATGTGGGGACTCGCCCCGGCCTTGCGGTGTTGACACTGTTTCACAAAAACAGCGAAGGGGGACCACTGGAGGCCAACGAGGGCCCTTAGTGGTCGCGTCTGGAACTACACCCGCTTCTT 248
H M W G L A P A L A C . H C F T K K T A K G D H W R P T R A P S G R V W K L H P L P
T C G D S P R P W R V D T V S Q K Q R R G T T G G Q R G P L V V A S G N Y T P R F L
A H V G T R P G L G V L T L F H K N S E G G P L E A N E G P . W S R L E T T P A S F
CACCAGAGAAGGGGTTAGCATTTTAGCATCCCGTAGCAGGACGATGCTAAACCGTCTTGAATGTACACATCCATGATCGAAGAAGCTCACGGCGAGAAGATGTTGATCGTTGGCTCCCA 248
S P E K G L A F . H P V A G R C . T V L E C T H P . S K K L T A R R C V D R L A S H
H Q R R G . H F S I P . Q D D A K P F L N V H I H D R R S S R R E D V L I V W L P
T R E G V S I L A S R S R T M L N R S . M Y T S M I E E A H G E K M C . S F G F P
CTGCGTGACTGTGCACGTTTTATTATCATCAGCCACGTTCTCCCCAGGTGGCCGGCTGTGGGGGATGCAGAAGAACCCTCCGGCCATGAACTACGACAAGCTCAGCCGCTCACTGCGCTATTAC 372
C V T V H V L F I S H V P P Q V A R L W G M Q K N R P A M N Y D K L S R S L R Y Y
T A . L C T F Y S S A T F L P R W P G C G G C R R T V R P . T T T S S A A H C A I T
L R D C A R F I H Q P R S S P G G P A V G D A E E P S G H E L R Q A Q P L T A L L
TACGAGAAGGGAATCATGAAAAGGTATGTTGAAGCTAAATGTCTGGCGTCTAATTAATGCCTCTGAATACAGCTCCTATACTAAAAACACTGAGGACTTTGATTGATGCCTGACTGAAGC 496
Y E K G I M Q K V C . S . M S G V L I K C L . I Q L L Y L K T L R T L I D A . L K
T R R E S C K R Y V E A K C L A F . L N A S E Y S S Y T . K H . G L . L M P D . S
L R E G N H A K G M L K L N V W R S N . M P L N T A P I L K N T E D F D . C L T E A
CACGCCCCCTTCTGCCCTCCTCAGGTGGCAGGCGCTACGTTTACAAGTTCGTGTGGAGCCGAGGCTCTCATCTCTGGCGTCCCCGACAATCAGCGGCCAGCCTGAAGGCCGAG 620
P R P L L A L S R W Q A S A T F T S S C V S P R L S S L W R S P T I S G P A . R P S
H A P F S P S P G G R R A L R L Q V R V . A R G S H L S G V P R Q S A A Q P E G R
T P P S R P L Q V A G E R Y V Y K F V C E P E A L I S L A F P D N Q R P S L K A E
TTGAGCGCTACGTCAACGAGGAGGACACCGTCCCTGTCCACCTGGACGAGGGCGTGTCTACACGACAGAACAAGCTCCACAAAACATGGGCCCCAGCCCTACTCCAAAGGCTACATGT 744
L S A T S T R R T P S P C P T W T R A C P T R Q N K L S H K T W A P S P T P K A T C
V . A L R Q R G H R P P V P P G R G R V L H D R T S T K H G P P A L L Q R L H V
F E R Y V N E E D T V P L S H L D E G V S Y T T E Q A P Q N M G P Q P Y S K G Y M
ACTAAAGCCCCCGACCCCTCTCTGCTTTCAGAACCCCTCTGCTGTACCCACCCACCATACACATCCCAAAGCCCCCGCCCTCCGCACATGCCCGCCCCACCCACTTGTATATAAGG 868
T K A P P T P L L L S E P L L L Y P P . T I I H I P K P R P L R T C P P H P L V Y K
L K P P R P L S C F Q N P S C C T H P P S Y T S Q S P A P S A H A R P T H L Y I R
Y . S P P D P S P A F R T P P A V P T H H H T H P K A P P P P H M P A P P T C I . G
CTATGGTATTTTTGTTTAAATTAATCTCGGTAGGATTTGATTGTTTTCCAGAGTTCGCTCCTCACATCTGAGGCCAACTCGACCCGTAACATGACTGTGGTCAAAGAGATGCTTAAT 992
A M V I F V L N . S R . G F D C F P E L R V L H D R I . G Q L D P . T . L W S K R L I
L W . F L F . I N L G R D L I V F Q S C V P L T S E A N S T R K H D C G Q R D A .
Y G D F C F K L I S V G I . L F S R A A F L S H L R P T R P V N M T V V K E M L N
AAATACCCATTATACTCAAAGTGTCTGTCATGAGTCTTATTACACTTCTTATTACAGTGTGTAGTCTGAAAGGCTAGCAGCAGGCTAGCAGCGCGGCTCAGCGGCGCTGAAGCGCGC 1116
N T H Y T Q T G V C M S S Y Y T S Y Y T C C S P E G . Q Q A S S A A S A A P E G A
. I P I I L K L V S A . V L I T L L I T R V V V L K A S S R L A A R R Q R R L K A R
K Y P L Y S N W C L H E F L L H F L L H V L . S . R L A A G . Q R G V S G A . R R

Exon 12
 Exon 13
 Putative Polyadenylation Signal
 Pufferfish HRH1 Homologue open reading frame containing the final exon

GGAGATCTCCAGCGTGGGCTCTCGTAGCGGTTGTAGAGCGGCTCCAGTTTCTGCTGCCTCTTCTGGCGGCTGAGCCGAGTCGGGTCGCCGACCCTGTAGAACGCCGGCGCAACTCCACC 1240
G D P P G V G L L V A V V E R L Q F L L P L L A A E P S R V A D P V E R R R E L H
E I L Q A L G S S . R L . S G S S F C C L F W R L S R V G S P T L . N A G A N S T
G R S S R R W A P R S G C R A A P V S A A S S G G . A E S G R R R P C R T P A R T P P
AGCCAGCGGGGTCGATGGTGGTCACCTCCCGCATGTACTCTTGGTGGTCAGCACCGCTCGTATAAACAGCCTGTGGAGACGGAGCCGTCACTGCACAAACATGGAGGTGCACCCGAGG 1364
Q P A G V D G G H L P H V L L G G Q H Q L V I N Q P V E T E P S L H K H G G A P A G
S Q R G S M V V T S R M Y S L V V S T S S . . T S L W R R S R H C T N M E V H P Q
A S G G R W W S P P A C T P W W S A P A R D K P A C G D G A V T A Q T W R C T R R
GTCATGTGACCCCGGCTCCAATCGGACTCACCCTCGGGCTGGCGGTTGAAGAGCGTGTGGAGGGGTGCAGGTACACCACCTGCTGGTCTATCAGGGTGCGGTACCCTGCTCGGGGTGCTTC 1488
S C D P G S N R T H H S G W R L K S V L E G C R Y T T C W S I R V R Y P S C G C F
G H V T P I G L T T R A G G . R A C W R G V G T P P A G L S G C G T R P A G A S
V M . P R L Q S D S P L G L A V E E R A G G V Q V H H L L V Y Q G A V P V L R V L
CTGGCGGCTTCTGAAGAACCCTGCAGATGGCTTTCTGAACCTGGACCGACGCTTTTTACAGGACACCAGTCCAGCTTGTGCCTGGGGAAGGTTTCCACGCAGAGTGTGAGCCGAGCC 1612
L A A F L K N P L Q M A F . T W T D A F S Q D T T S S L C L G K V S T Q S V S R R
W R R S . R T R C R W L S E P G P T L F H R T P R P A C A W G R F P R R V . A D A
P G G V P E E P A A D G F L N L D R R F F T G H H V Q L V P G E G F H A E C E P T P
CCCCATGACCTCGCTCAGGCTAGCTCCGCTCTCACCTGTCCATGCTCAGCATGTGCTTGGGATGCTCCTGAGCCCTCTCAGGGAGCGCCCTGGATGAAGTTCTCGAAACACCAGG 1736
P P . R S L R L A P P P H L S M M L S I C L R M S . A L F R E R A W M K F S K H Q A
P H D A R S G . L R L L T C P . C S A S A C G C P E P S S G S A P G . S S R N T R
P M T L A Q A S S A S S P V H D A Q H L L A D V L S P L Q G A R L D E V L E T P G
GTTGGAGAACTTGTGTCTTCCAGGAGTGTAGACGGCCAGCAGCGTCAGGTGGTCCCTCCAGCTGGAAGAACTTTGTCTTCTTGTGTCGGCGAGAGCCTGTTGTCTGCTGGAGACATCAA 1860
L E N L L F F Q E L . T A S S V R W S P S S W K N F V F F W S A R A C L S W R H Q
R W R T C C S S R S C R R P A A S G G P P P A G R T L S S S G R R E P V C P G D I K
V G E L V V L P G V V D G Q Q R Q V V P L Q L E E L C L L L V G E S L F V L E T S
ACATCATCATCATCTTAACTCCACAAAAGATCAATAAAGAACCAGCGAGAGCTTCAAATGCCGATCGATAAAATCAGGGCTGACGTGAGAACGCTTCTGCTGCTCCAGAACCC 1984
T S S S S S S . T P P K R S I K N Q R E L Q M R I D K I R A D V R T L P A A P R T
H H H H H L K L H Q K D Q . R T S E S F K C A S I K S G L T S E R F L L L P E P
N I I I I I L N S T K K I N K E P A R A S N A H R . N Q G . R Q N A S C C S Q N P
AACGCTGGTCACCATGGGAACCAGCGAGCTGGAGCGGGGGCGACATCGCACCTTTGGTCTGTAGAAGATGTTCTGCACCCGACAGCATGGAGACGATGGTCAGCATCTCGCTGCAGCCCA 2108
Q R W S P W E P G E L E R G R H R T F G L . K M F C T R Q H G D D G Q H L V A A A Q
N A G H H G N Q A S W S G G D I A P L V C R R C S A P D S M E T M V S I S S L Q P
T L V T M G T R R A G A G A T S H L W S V E D V L H P T A W R R W S A S R R C S P
GATGGACGGACATGATCAACATCTTGACAGCATGGGCTCCAGGGGGAACCTGCCATCTGAAACACACACACCGCCAGCGCACGTGCGCAGGTGCTGTGAGTGCTCAGCGGGTCCCGG 2232
M D G H D Q H L A Q H G L Q G E L C H L K H T H T R T R T C A G R C E C S A R S R
R W T D M I N I L H S M G S R G S A I . N T H T R A R A R A Q V A V S A Q R G P
D G R T . S T S C T A W A P G G T L P S E T H T H A H A H V R R S L . V L S A V P

_____ Pufferfish HRH1 Homologue open reading frame
containing the final exon

ACGGTCACGCCAGACACCCAGCACGCTCACCCCTCTCCCCAGCCGCGTGAGCAGCCCTCGTCGTCAGAGCCCCAGCGGTAGAGCTGCTCCATGGCCGTGATCAGCGTCTCCATGGGGGG 2356
T V T P D T P A R S P S S P A A . A A P R R P E P P A C R A A P W P . S A S P W G
R S R Q T P Q H A H P P P Q P R E Q P L V V Q S P Q R V E L L H G R D Q R L H G G
D G H A R H P S T L T L L P S R V S S P S S S R A P S V . S C S M A V I S V S M G G
GGAGTCCATGAAGTCGAAGGACAGCAGGTCGTTGACGCCCATCGCCTGACGGGGGAGACCAACACCTCTGAGCACGTTCAGGAGGGGGGGCTTCTGGTAGGAGGGGGG 2467
G S P . S R R T A G R . R P S P D G G D Q T P L S T F R R G G V L V G G G
G V H E V E G Q Q V V D A H R L T G E T K H L . A R S G G G A F W . E G G
E S M K S K D S R S L T P I A . R G R P N T S E H V Q E G G R S G R R G G

Appendix F Pufferfish ERM Genomic Sequence

The genomic sequence of pufferfish ERM is given. The deduced amino acid sequences that bear homology with other ERM protein sequence and that correspond to open reading frames in the genomic sequence are boxed. Exon numbers have been indicated throughout the sequence. The position of a putative start codon is highlighted in green and a putative polyadenylation signal is boxed in red.

CAGGTTTGGTGCCAGTGTAAATGTGTAGGTTAGCATGGACGGGCAGCATACCCACGGCACAAACAGGTACATTTGCGATTGTATTTTTAGCGTGAAGGAAAAATTTGTTAATTATATTTACATA 124
Q V W C Q C K C V G . H G R A A Y P R H K Q V H L R L Y F . R G R K I C . L Y L H
R F G A S V N V . V S M D G Q H T H G T N R Y I C D C I F S V E G K F V N Y I Y I
T G L V P V . M C R L A W T G S I P T A Q T G T F A I V F L A W K E N L L I I F T Y
CTTCATCACCACGAACTGTGGAGAAATGGGTGCAATGTTGGTGCTAAATCTATTTCAAAAGAGGGACTGGTTATAATTCAACTGTGGCTAAAATGGTTGCTTATAATAAGAAAGAGGTTTAA 248
T S S P R T V E N W V Q M L V L N L F Q K R D W L . F N C G . N G C L . . E R G L K
L H H H E L W R I G C K C W C . I Y F K R G T G Y N S T V A K M V A Y N K K E V .
F I T T N C G E L G A N V G A K S I S K E G L V I I Q L W L K W L L I I R K R F E
ATGTACTAATGATATTGATACATAGATTAATGATAGAATATAAGGAAACCAGGAATAACAAAATCTATAGTTGGATAGAACAAATTAAGGCATAATACAGCTGACAGTATTATCTCCAGGTG 372
C T N D I D T . I N . . N I R K P G I T K F Y S W I E Q I K A . Y S . Q Y Y L Q V
N V L M I L I H R L I D R I . G N Q E . Q N S I V G . N K L R H N T A D S I I S R W
M Y . . Y . Y I D . L I E Y K E T R N N K I L . L D R T N . G I I Q L T V L S P G
GCCTATCACCAGTAACATAAAAATACTGCTATGTGTATGTTGGCATAAAAATGCTCTTTGTTGATTATTGCACTTTTTGTCTTTAATGCTTTTGACTGTCAACAACAACAAGAACATTC 496
A Y L P V T . N T A M C M L C I K L C S L L Y Y C T F C L . C F . L S T T N K N I
P I Y Q . H K I L L C V C C A . N C A L C C I I A L F V F N A F D C Q Q Q T R T F
G L S T S N I K Y C Y V Y V V H K T V L F V V L L H F L S L M L L T V N N K Q E H S
TTTGATACAAATCAGTCACACAGTAAAGTTTATCACAGTTGCATAGTTTACGATTTTTAGCTTCATGCAACTAGAAGCCAAATAAGCCAAAAAAGGAGTATTGAAAATAGTCACTAA 620
L C I Q I S H T Q . S L S Q L S H S I V S T I F S F M Q L E A Q I S Q K K G V L K I V T N
F V Y K S V T H S K V Y H S C I V S R F L A S C N . K P K . A K K K E Y . K . S L
L Y T N Q S H T V K F I T V A . F H D F . L H A T R S P N K P K K R S I E N S H .
TAAGCTCTGTAGATTTCTAATTATCGGGCTGGCAGCATAATGTAGAGAATTTTCATAAATCTCAGAATGGGTTTTGTTGAAAACATGGGCAACAATGTTTACAGAATTTTTAGCTCGAGCAT 744
K L C R F L I I G V A A . C R E F S . I S E W V L L K T W A T M F Y R T F . L E H
I S S V D F . L S A W Q H N V E N F H K S Q N G F C . K H G Q Q C F T E L F S S S I
. A L . I S N Y R R G S I M . R I F I N L R M G F V E N M G N N V L Q N F L A R A
CTAATAGTATGAATTAGTTTTGGTGAATTTGTTGGCAGTACAGAAAAAAAATCATTGTACCATTACTGCTAACTGGTTTTCATCGGTTATTAGTTCAGGGAGACAGCAGCTGCCTACTGACCA 868
L I V . I S F G G I C L A V Q K K K S L Y H Y C . L V S C V I S S G R Q H C L L T
. . Y E L V L V E F V W Q Y R K K N H C T I T A N W F H A L L V Q G D S T A Y . P
S N S M N . F W W N L F G S T E K K I I V P L L L T G F M R Y . F R E T A L P T D Q
ATCAATGATCTGAAAAATGACATCATTAGAGCAGATCCCTGAGCTCGGATGGGAGAGGTGTGCTGCAGACATGAAAAGGAAAGGAAAGTAAAATGAAAAATAGAGAGCATATCTAAAACAA 992
N Q . S G K . H H S E Q I P E L S L G W E R C A A D M . K E R K . N W K I R E H I . N N
I N D L E N D I I Q S R S L S S D G R G V L Q T C K R K G S K T G K L E S I S K T
S M I W K M T S F R A D P . A R M G E V C C R H V K G K E V K L E N . R A Y L K Q
CATAATGGAGCTACTTTTTGTTCTGCATATCCACACAGCATGGATGCCATCTGTACTGGCTTCTCTCTCAAGAAATGAAAGTCTCTCTGAGCTCGTATTCAAGTCAAGACGTTGCGTAAATGC 1116
I M E L L F V L H I H T A W M P S V L A S L L K K . K C S C S S Y S S Q D V A . C
T . W S Y F L F C I S T Q H G C H L Y W L L F S R N E S A P A A R I Q V K T L R N A
H N G A T F C S A Y P H S M D A I C T G F S S Q E M K V L L Q L V F K S R R C V M

TGTGATTGATGTCAGAGTTTTTATAGCCACTGTAATGCCTTTTGTGGGGTTTTGAAAATACAAACAGCAAGATACACAAAAACATCATAGCCTGGGAATCCCAGAAATCATGGTAATAGATT 1240
 C D . C Q S F Y S H C N A F C G V L E I Q T A R Y T K H I I A W E F P E I M V I D
 V I D V R V F I A T V M P F V G F . K Y K Q Q D T K N T S . P G N S Q K S W . . I
 L . L M S E F L . P L . C L L W G F R N T N S K I H K T H H S L G I P R N H G N R L
 AAACATACAGATATTTTGTAGGGTGAAGACTCTACGATATGGATGGCTTCTATGACCAGCAGGTTCCATTGTAGCTTCGGAGAGTGTGAGTTTCAAAGATTTCAAGCACATATGCAATCTT 1364
 . T Y R Y F C R V K T P T I W M A S M T S R F H L . L R R V . V S K I S S T Y A I L
 K H T D I F V G . R L L R Y G W L L . P A G S I C S F G E C E F Q R F Q A H M Q S
 N I Q I F L . G E D S Y D M D G F Y D Q Q V P F V A S E S V S F K D F K H I C N L
 GGAAAAGAGCTGATTGTGACAGTTTTGCTCTTCAAACCTGTAGAAATGGCCCTCAGAAGTTGCTGAGAAATGCCAGCTAGGCGGAAGAAGAAGTTTCATGGACACAGAGTTGGCCAGGAC 1488
E K S . L C D Q F S L F K L . K W P S E V A E K C P A R R K K K F M D T E L A Q D
 W K R A D C V T S F R S S N C R N G P Q K L L R N V Q L G G R R S S W T Q S W P R T
 G K E L I V . P V F A L Q T V E M A L R S C . E M S S . A E E E V H G H R V G P G
 ACGGAAGGTAGCTGCACCATCTATGATGACATGAAACAATGATACAATGCAATAGATGCCAGTTAATAACTGAAGTACCAATGTGGTTATAGCATGTGAAACCACCCAGATGGAAAAAGACAC 1612
T E G S C T I Y D D M K Q . Y N A I D A S . . L K Y Q C G Y S M . N H P D G K K T
 R K V A A P S M M T . N N D T M Q . M P V N N . S T N V V I A C E T T Q M E K R H
 H G R . L H H L . . H E T M I Q C N R C Q L I T E V P M W L . H V K P P R W K K R T
 TGGGACAATTATTTTTGGCTGTAATTGAATACAACCTGATTTCAAATGAATTAGGATAATGATGGACGTCCTGTGATGATTACCAGAAATTAAGGCCAAGATGGTATTTCTGATGACG 1736
 L G Q L F F W L . L N T T . F Q N E L G . . W T S L . . I H R N . G Q E W Y F . . R
 W D N Y F F G C N . I Q P D F K M N . D N D G R H C D R F T E I K A K N G I S D D
 G T I I F L A V I E Y N L I S K . I R I M M D V T V I D S P K L R P R M V F L M T
 TGTGTTTTAAAAAGACTCTTGACAGCTTAAATCTTAATGTCGTTTCATCACTGTTTTGACAGCTCTTTCAAGATTTAAGTACAGCTTCAAGAAATCTGGATAGCAGAAGGTGAATTTTGTG 1860
 V F . K . L L Y S L N L N V V S S L F C R A L S R F K S A S R N L D S R R . I L L
 V C F K N D S C T A . I L M S F H H C F A E L L F Q D L S Q L Q E I W I A E G E F C .
 C V L K M T L V Q L K S . C R F I T V L Q S S F K I . V S F K K S G . Q K V N F A
 AATAAACAGCATTACATGTTAGTTTTAGTCTTCTGTAGGCCACTGGCAGGAATGATATTGTTGGTGTCTGGAAATGTTGCAGTACCAAATGTGGCCTCTGGGTAACCAACAGCAA 1984
 N K T A F T C . F . S S L . A H W Q E . Y C W C L W K C C S T K C G L W V T Q Q Q
 I K Q H L H V S F S L L C R P T G R N D I V G V C G N V A V P N V A S G . P N S K
 E . N S I Y M L V L V F S V G P L A G M I L L V S V E M L Q Y Q M W P L G N P T A N
 CTTGTGACAGCTTCAAGTAAAGCCTCAAATTCAGTAGCCAAGTGACCAAGTGCAGAGTGCTAAAGGCCAGTTGGGAGATTGTTGGCTGCACCACGATTTAAAAAGTCTTTTTCTCTT 2108
 T C . H A S Q V K P Q N S V A K . P E C . R A S W G D C L A A P R F . K V L F F L L
 L V D T L L K . S L K I Q . P S D Q S A K G P V G E I V W L H H D F K K F F F S F
 L L T R F S S K A S K F S S Q V T R V L K G Q L G R L F G C T T I L K S S F F P S
 CACAGCTCAAGTTCCTGATGATGAGCAGTTTGTCCCAGATTTTTCAGTCCAACAACCTGTAAGTAGTAAGCAAAGTTGTGTAATTTCTTCATTGTAACAAAAACCACTGTAATTTGTAAGTAAT 2232
T A Q V P D D E Q F V P D F Q S N N C K . . A K L C K F F I V T K N H C N L . S N
 S Q L K S L M M S S L S Q I F S P T T V S S K Q S C V N S S L . L K T T V I C K V M
 H S S S S . . A V C P R F S V Q Q L . V V S K V V . I L H C N . K P L . F V K .



Exon 2



Exon 3



Exon 4




Exon 5



Putative Start Codon

GTTTCCAATTTCCAGGTTTTCTCTCGGTTAAATGTACAACCCATAAATCTGGCTATGCTGTGCTGCTATATATTTTGACTAAATTAATGTATGCTGCAATAAGTGACAGCAAGAACCAAAAT 2356
V S N F Q V F L L V K C T T L K S G Y A V L L L Y I L T K L N V C C N K . Q Q R T K
F P I S R F F S W L N V Q P . N L A M L L C C Y I F . L N . M Y A A I S D S K E P N
C F Q F P G F S P G . M Y N P K I W L L C C A A I Y F D . I K C M L Q . V T A K N Q I
ATTTGGAAGCACAGTGCTTAATAATACATCACATGTCCTCAACAACAACCTCTGATATCGAGTGCCTACATTACCAGGTACAGTGTGACTGGCAGCCATATAAAGATATGAGCCACACCCTGG 2480
Y F E A Q C L N T S H V P N N N S D I E C V H Y Q V T V F D W Q P Y K D M S H T T G
I S K H S A . I H H M S P T T T L I S S A Y I T R S Q C L T G S H I K I . A T P L
F R S T V L K Y I T C P Q Q Q L . Y R V R T L P G H S V . L A A I . R Y E P H H W
TGGCAGGTAATGTTGGATTATAATGCATCGACGTCATAGATGTTTCATGGAAGGACATTTACAGTTAACGCTGCTGTGGTGGGGCTTTAAGATCAGAAGCTCAATATGGCCCAATGTTATTAG 2604
G R . M L D L . C I D V I D V H G R T F T V N A A V V G A L R S E L N M A Q C Y .
V A G K C W I Y N A S T S . M F M E G H L Q L T L L W L G L . D Q N S I W P N V I S
W Q V N V G F I M H R R H R C S W K D I Y S . R C C G W G F K I R T Q Y G P M L L
TTACTGAATAAATTTACTGGATGAAAAGAGTAAGTTCCTCCATAAGCAGTGGTGGAGTGTGCATTAGAGTCAACAAGAGATCTTTAAATGTCAGTCTACAAAGGTATTAGCGAAAAGGAGTTTTTCAT 2728
L L N K F T G . K E . V P I S S G G V C I R V N K R S L N V S L Q R Y . R K E F S
Y . I N L L D E K S K F P . A V V E C A L E S T R D L . M S V Y K G I S E R S F H
V T E . I Y W M K R V S S H K Q W S V H . S Q Q E I F K C Q S T K V L A K G V F M
GAAAATCAGGTTTGTCCAAATATTATAAAAACCTCCACTTACATTTCCACATTTTTCGACAGACTTTTTTGAGACCCTCTCTCTGTTCAGGGTTAGGGTTTTCTTTTCTGCGAAAATGCATTTAT 2852
. K S G L F Q I L . N S H L H S T F L Q T F L R P L S L F R V R V S L S A K M H F I
E N Q V C S K Y Y K T P T Y I P H F C R L F . D H S L C S G L G F L F L P K C I L
K I R F V P N I I K L P L T F H I S A D F F E T T L S V Q G . G F S F C Q N A F Y
TTATTTGTATTTATTGAAGTATTTATAGGGATAAGATCAAGACAATTCGGCTTCTCTGCTCACATGTGCCTGGGATAGATGCACAGCTGTATTACTTTAACATATGACCTCTGACCCCTT 2976
Y F V F I E V F Y R D K I K T I P A S L L T C A W D R C T A V F T L T Y D L . P L
F I L Y L L K Y F I G I R S R Q F R L L C S H V P G I D A Q L Y L L . H M T S D P L
L F C I Y . S I L . G . D Q D N S G F S A H M C L G . M H S C I Y F N I . P L T P
ATGAAAAATCCCTTCATCCTTTCTCAGTGTAAGGACAGATAAGAAGCCTTTGTTGTTCTTTGATGCTGCTGCTGACCAGTACGGAGCCATAAACGTAACAGCCACGCTCTTTTCATCA 3100
M K N P F I L S S V . R T D K K P L F V L . C S A R H Q Y G A I N V T A T S L F I
. K I P S S F L Q C K G Q I R S L C L F F D A L L V T S T E P . T . Q P R L F S S
Y E K S L H P F F S V K D R . E A F V C S L M L C S S P V R S H K R N S H V S F H H
CCTCCTACTTGAGGAAAACACAGAGCTTGATGTTGGAAGTTTATGTAGAACAGAATTTAGATTTTCATGTGCACAAGTACATGACATGTAATCAACAGCTAGAACACAAAAATGAATTTT 3224
T S Y L R L L K Y F I D V G S F M . K Q N L D F M C T S H L T C N Q Q L E H K N E F F
P P T . G K L Q S L M L E V L C R N R I . I S C A Q V I . H V I N S . N T K M N F
L L L E E N Y R A . C W K F Y V E T E F R F H V H K S S D M . S T A R T Q K . I F
CTAGTGCAAAGATGTAATTTATAATAATATATTGGCAGTGATGTTTCATGGACCGCTGCTCAATAAAGTGAAGCAAGAGTTCAGTGCCTTCTGCAACCTCTCATGTTAATGCGTGCCTT 3348
. C K D V I I N N I L A V M F H G P P V N K V K Q E F S A S S C T T S H V N A C L
S S A K M . L L I I Y W Q . C F M D R L S I K . S K S S V R L P A Q P L M L M R A F
L V Q R C N Y . . Y I G S D V S W T A C Q . S E A R V Q C V F L H N L S C . C V P

 Exon 6

ACACCTGCCAGTCATCATCATTCCGGGAAGTCTCCAGTCTTGTACCCAGAATAACAGCAGGGGTGAGTCCCTTGTGGCCCTTTGAGCCTGTTTTAATGGGAAAAACATGATCTTTTCTAATTCCT 4588
 H T C Q S S S F G K S P V L Y Q N N S R G E S L L A F E P V L M G K T . S F L I P
 T P A S H H S G S L Q S C T R I T A G V S P C W P L S L F . W E K H D L F . F P
 S H L P V I I I R E V S S L V P E . Q Q G . V L V G L . A C F N G K N M I F S N S L
 TCCTTTAAACCTCCACAGGCTTCGTCCTGACAGGGATCCTTGTTTAATGATGATGATGCTTGTGTGGCCAGAAAGATTGGAAGGTAATCATCCAATGAAGGTTTTAAAGGGTGGGAATG 4712
 F L . T S T G F R P D R D P C L M Y D D A C V V P E R L E G K S S N E G F K G W E W
 S F K P P Q A F V L T G I L V . C M M M L V L C Q K D W K V N H P M K V L K G G N
 P L N L H R L S S . Q G S L F N V . . C L C C A R K I G R . I I Q . R F . R V G M
 GGGGACATACTGTAGACTGTGACTCAATGTGTGTATGTGTCAGCAGGTAAGGGGAAGCAGGACTGTTTGCCATACCAGCGCCGGGGCTCCTTGACACTCTGGCAGTTCTTGCTCACCCCTC 4836
 G T Y C S T V D S M C C M C Q Q V K G S R T V C H T S A G A P C S S G S S C S P F
 G G H T V A L W T Q C V V C V S R . R E A G L F A I P A P G L L A A L A V L A H P S
 G D I L . H C G L N V L Y V S A G K G K Q D C L P Y Q R R G S L Q L W Q F L L T L
 TGGACAACCTTCAAATGGGCACCTGATCATTGGACTGGCGAAACATGGAGTTAAACTCATTGATCCCGAAGAGGTGAGAGAATGAACCTTGACCGTATTAATTTGTAGAACAGGGATTC 4960
 W T T L Q M G T . S F G L G E T W S L N S L I P K R . E N E L D R I K L C R T G I
 G Q P F K W A P D H L D W A K H G V . T H . S R R G E R M N L T V L N Y V E Q G F
 L D N P S N G H L I I W T G R N M E F K L I D P E E V R E . T . P Y . I M . N R D S
 CGGACTAACAGAGAAGAACAGCAGCTTTAAAGATGAATGTATCAATGTGAAATAGATTTGAATTC AATAACGCTCTTAAAAATTGATGAAATTTCTTTTGCAGATGTGAAATGTATGCTTTTT 5084
 P D . Q R R T A A L K M N V S M . N R F E F Q . R L K N . . N F F C R C V N V C F L
 R T N R E E Q Q L . R . M Y Q C E I D L N S N N V L K I D E I S F A D V . M Y A F
 G L T E K N S S F K D E C I N V K . I . I P I T S . K L M K F L L Q M C K C M L F
 ATTAATGCCATTTATTAAGGTGGCTCGGCTCGGGTGTCCAAAAGAAATCGACCCGCCATGAACATGACAAGCTGAGCCGATCGCTGCGATATTACTACGAAAAGGGCATAATGCACAAGGTA 5208
 L M P F I K V A R L W G V Q K N R P A M N Y D K L S R S L R Y Y Y E K G I M H K V
 Y . C H L L R W L G S G V S K R I D R P . T M T S . A D R C D I T T K R A . C T R .
 I N A I Y . G G S A L G C P K E S T G H E L . Q A E P I A A I L L R K G H N A Q G
 ATGGCGGATAAAGATATTTACTGGATTAGAGCTGTTGTCTGATTTACAGGTCTAATTTTCATGTGGCCCTAAAAGGGATCAAGCGTTCAAGAAGATGAAATAAAAAAATGAATCCCAACTGA 5332
 M A D K D I Y W I R A V V . F T G L I F M W P . K G S S V Q E D E I K K . I P N L
 W R I K I F T G L E L L S D L Q V . F S C G P K R D Q A F K K M K . K N E F P T .
 N G G . R Y L L D . S C C L I Y R S N F H V A L K G I K R S R R . N K K M N S Q T L E
 AACATCCCTGGGTTGAGCCGCTCATTAAAGGCTTTGTTAATCACTGTCAGCAGTTTACATTTTATAATGCGGATTAAGATTGCCTCACAACTGTTCCGATTTATTCCG 5456
 K H S L G L R P C L G L I K G F V N I T L Q Q F Y I L . C G L K I A S Q S F R F I R
 N I P W V . G R A W V S L R A L L T S L C S S F T F Y N A D . R L P H N R S D L F
 T F P G F E A V L G S H . G L C . H H S A A V L H F I M R I K D C C L T I V P I Y S
 ATGAATATTGTTAGTAAATGTCTGTGCTTCCTATCATCCTGCAGGTGTGCTGGTGAACGGTACGCTACAAAGTTCTGTAATCTGGAGGCATTTTCTCCATGGCTTTCCAGACAACCAGAG 5580
 . I L L V N V C A S Y H P A G C W . T V R L Q V R L . S G G T F L H G F P R Q P E
 D E Y C . L M S V L P I I L Q V A G E R Y V Y K F V C N L E A L F S M A F P D N Q R
 M N I V S . C L C F L S S C R L L V N G T V S T S S S V I W R H F S P W L S Q T T R

Exon 10 Exon 11 Exon 12 Exon 13

GCCAAGTCTGAAGAGTGACGTGGATGCAATCCTGCCTCCGAGTGAGGACGATGGCTTCCCACCTACTTATGAGGAAGAGGGTCCGTATCTGACCGAAGGAGGAGAACAGTTTCACAGTCT 5704
A K S E E . R G C N P A S E . G R W L P H T Y L . G R G S V S D R R R R T V H P V
P S L K S D V D A I L P P S E D D G F P I P T Y E E E G P Y L T E G G E Q F I Q S
G Q V . R V T W M Q S C L R V R T M A S P Y L L M R K R V R I . P K E E N S S S S L
TTTCCCACAGCTATCCTTACTAGATCCCACGGCCTTAATTACTGGGCAGAAGTGCATATGAGATGTACGCCTGCATTATATATGCCACAGAATAGATAAATTATTATTTTAAAGTCTACTGG 5828
F S R Q L S L L D P T A L I T G Q N C I . D V R L H Y I C H R I D K L L F F K S T G
F P D S Y P Y . I P R P . L L G R T A Y E M Y A C I I Y A T E . I N Y Y F L S L L
F P T A I L T R S H G L N Y W A E L H M R C T P A L Y M P Q N R . I I I F . V Y W
AAGTTTTGTTCTAAAAGTTTAGTTTAAATAAAAA 5863
S F V L K V . F K . K
E V L F . K F S L N K K
K F C S K S L V . I K N

Exon 13 Putative Polyadenylation Signal

Appendix G Two-way Alignment of the Mouse and Pufferfish PEA3 Promoters

The two-way alignment of the mouse and pufferfish PEA3 promoters is illustrated. The position of known exons within each sequence has been depicted by boxes on the sequence with the exon number indicated. The start site of transcription has been shown (+1) in the mouse sequence. The start site of translation has been indicated in each of the sequences.

Mouse PEA3 Promoter G A - A T T C T G G G G T C A G A C T G A G T T G A G A G A C C C T G T C T C A A A A A C A T G G T A
 Pufferfish PEA3 Promoter G G T A C T T T G G A T C G G A C C - - - - - G C C T C - - - - - T G G - -

Mouse PEA3 Promoter A G G T A A G T A G G A A G G A A G G A A G G A A G G A A G G A A G G A A G G A A G G A A G G A A G
 Pufferfish PEA3 Promoter - - - - -

Mouse PEA3 Promoter G A A G G A A G C C A C A A C A A A C C A T G A G C T G A G A G A C G G T G C A T C A G T C T A G A
 Pufferfish PEA3 Promoter - - - - - A G C T G C A G C - - - - - C T - - - - - C T G G A

Mouse PEA3 Promoter G C A C T T G C T G C T T T T T C A G A G G A C C T G A G T T C T G T T C C C A G C A C C T A C A T
 Pufferfish PEA3 Promoter G - - - - T T G C A G C C T C T - - - - - G G A G C T G - - - - - C A G C C T C T - - - -

Mouse PEA3 Promoter C A G A C G G C T C A C G A C T G C C T G T T A C C T G C C G C A T C A G G A A A T C C A G T G T C
 Pufferfish PEA3 Promoter - - - - - G G A G C T T T A G A T C C

Mouse PEA3 Promoter T C T G G C C T C T G T A G G T A C C C A C A C A A C C T A C A C A A A C A C A T A A A T A A A A
 Pufferfish PEA3 Promoter T G C G G T T T C C G T G T G - - - - - G C G C A G A C C T A - - - - - A A T A A A A

Mouse PEA3 Promoter A T T A A A T T A A A A A T A A A T A A C G C A A A A T A A G A A A A A A A C G T T A T A A A A A
 Pufferfish PEA3 Promoter A - - - - A C C T C A G G A A T - - - - - T - - - - - C A A G A A

Mouse PEA3 Promoter T T G T G T G G G G A G A G T C A G C C T C T T T G A T G G C T T C A G A G C C A G C T T G C C A
 Pufferfish PEA3 Promoter T T - - - - - C T T C A G G A - - - - - T T G C - -

Mouse PEA3 Promoter G T T A C A T T C A G G C T T T C T T C T T C A A G C T G T A T C A G G G A T C T C C C A G C
 Pufferfish PEA3 Promoter G T C A - G T T T G G G - - - - - A G C T G A T A T A A G G G C T - - - - - G C

Mouse PEA3 Promoter T C C C A G G G C A A A C A G T G G G T T C C C T T T G T A T A A A A T T T A C A T A T C T A A A
 Pufferfish PEA3 Promoter T T T T - - - - - T T T T T C T T C T A T - - - - - T T T G T C T G T - - - -

Mouse PEA3 Promoter G T C C A C A T T T G C A T T T T A T A A T T T A A T T C C T A G G A A A C A G G T T A G A G T C
 Pufferfish PEA3 Promoter - - - - - T T T T C A T T T T A A A A C - - - - - A A C A C T - - - - - G A C G C G A G T C A A A G T T

Mouse PEA3 Promoter C A G C C C C C C A A T C T A C T G G A G G A A G A G A A T A T T T A C C T G C C C A G C T T C C C
 Pufferfish PEA3 Promoter A A A - - - - A C A A - - - - A C A A A A A A A A A A A A - - - - -

Mouse PEA3 Promoter G C G C T T C A T A A A T G C T A G T T A T T A T G G T T C C T C T G G G C G C A C A A G T C C G T
 Pufferfish PEA3 Promoter G C - C T C C A C G - - - - - C A G C T C C A T T G A - C G C G C A G - - - - -

Mouse PEA3 Promoter T T G C C T C C C C C T C C C C A G G C A C A C A G A C G T G T T C A C A C T C C C A C G A G A
 Pufferfish PEA3 Promoter - - - - - C T C C C T C C T C T C A - - - - - T C T G A G T C C C T G C G A - -

Mouse PEA3 Promoter T A G A T G C A C A G A G G C G T G C T G A A C G C C C G C G C A A G G A G C T C A G T C C T T A
 Pufferfish PEA3 Promoter - - G C T G C G C G G A - - - - - C C - - - - - A G C T C A G C C C C G A T

Exon 2

Mouse PEA3 Promoter G A C C C A A A A C C C A G G T T G A A C C C G T G G A G A A G C T G C C G G G T C T T C A G C T
 Pufferfish PEA3 Promoter G A C C A C A A A C - - - G G T T - - A A C - - - G G A - - - - - T G G A T T A T A A G - -

Putative In-frame Start Codons

Mouse PEA3 Promoter T C C G T C T T T T T T C T C A A T T C T T T A T G A A T G A A C T A A A G C C G A A A G G A G T
 Pufferfish PEA3 Promoter - - - - - A T G A T G G A - - - - -

Mouse PEA3 Promoter T T A T T T T T T A T G A A T G G A A G T C C A A A A A G T G A A T G A A G C C A G G A G C C A G
 Pufferfish PEA3 Promoter - T A T T T - - - - - G G A - - - C C A G C A A G T G - - - - -

Mouse PEA3 Promoter C C C C T A C T T T C T G A T T G G T C T T C G T G G C T G G G A A A C T C C T C C C T C A A G C T
 Pufferfish PEA3 Promoter - C C T T A C A C T T T A G C A A A T G T A C G T A - - T G G A - - - - T G C - - - - - A G C T

Mouse PEA3 Promoter C A G C T A G C C G A G T C T G A G G A T T G G C G G G C G T G C C C G G C C T G G G G T T C A
 Pufferfish PEA3 Promoter C G A C C G T C C G - - - - - G C T C G G C A C - - - G G C T T T A

Mouse PEA3 Promoter T T C A C A A A A G T C A A T G A A C A A A A G G A A C C G G T G G G A G G G C G G G G A
 Pufferfish PEA3 Promoter T T - - - - -

Mouse PEA3 Promoter A G A G C G A G A A C G G A G G C C A A G G C A A A G G A A A T G C A C C A A T C A G C T G C T C C
 Pufferfish PEA3 Promoter - - - - - T G T T T T A A T C - G T C G C - - -

+1 Exon 1

Mouse PEA3 Promoter C C C G G G C T C A C A A C T G T C T G C T G C G C C C G G A A A A C A A G C C G G G C T C T G G
 Pufferfish PEA3 Promoter - - - - - G T C - G C - G C G T C C G - - - - - A G C - - G A G T T T G -

Mouse PEA3 Promoter G G A C C C G G G C T C A G G C C G C C T C G C T C C G G C C T A G C C C C G C C A C C T T A G T
 Pufferfish PEA3 Promoter - - - - - C T C C G T G G C A G C - - A G C C C T G - - - - -

Mouse PEA3 Promoter TGTGT CATCCCCCGGGCATGCTGAGCATCCCCCGCGGGCTCCGGCACAGA
Pufferfish PEA3 Promoter TGTGT - - - - - GTGCGCG - TGTGTGTGCGCGCGCG - - - - - AGA

Mouse PEA3 Promoter CGCCCGGACCTCA GGTGAAGCTGCGGAAGGGGAAGGAAGGAGAGTGTCT
Pufferfish PEA3 Promoter - - - - - GCGAGGTGTGCG - - - - - TGTGT

Mouse PEA3 Promoter GTGATTTCCCGCTTCACTCCACACACCACTCTCTCTGTTCGGCGC
Pufferfish PEA3 Promoter GTG - - - - - TTTCTGTTCGA

Mouse PEA3 Promoter CCGCTTCCGCACATTCGGGGCGGGGGTGGGAAGCGTCGGGATTCGAGGC
Pufferfish PEA3 Promoter C - - - - - GAACTCGTC - - - - - C

Mouse PEA3 Promoter TGGAGGGGGCGCTGAGTTATGGGCGTGC GAA TGGTGTGCGCCGCGCGGC
Pufferfish PEA3 Promoter TAA A - - - - - CTAGT - - - - - GAA - AA TGT - - - - - T

Mouse PEA3 Promoter GCGCCCACTAACA CACACACACACATACACA CGGGACACACACACGCACA
Pufferfish PEA3 Promoter GCGCTAAGT - - - - - CAATCACGCGC - - - - - CGGG - - - - -

Exon 1'

Mouse PEA3 Promoter CACAC TTTATGTAAC CAGTTGGGCTGAGCAGGGCTGCAGAAAGCAGACAC
Pufferfish PEA3 Promoter - - - - - TTTG - GTGA T - GT TTTGGG - - GAGCCGAGC - - - - -

Mouse PEA3 Promoter ACCAACC CAGCTCAAGGGAGCAG GTGG GACC TCTTT TCGGGAGAGGAGT
Pufferfish PEA3 Promoter - - - - - GGC CA CTTT - - - - -

Mouse PEA3 Promoter GCGGTGGAGCGGCTCGG CCGCGCAGGTCCCA GCGAGTCCGGAGAAGTTTC
Pufferfish PEA3 Promoter - - - - - CCT CGCAGGC - - - - - GCGCGTCC - - - - -

Mouse PEA3 Promoter GC TGT CCGATCCAAAGCCTGGNGA TCGCCCTCTCCCCCGCGCAGCGCTC
Pufferfish PEA3 Promoter - - TGT T - - - - - T TGC TTTT T TCCCCGCGCGGA - - - - -

Exon 2

Mouse PEA3 Promoter CCGGGAACCGCGCGGAACTGCTCC CCCCCTCCAGG TCTCTGCCCTCCGCG
Pufferfish PEA3 Promoter - - - - - A CCGGAAACGTGTCT - - - - - GCTTC - - - - -

In-frame Start Codons

Mouse PEA3 Promoter CGGGGGCCCGGCCCTGTGGCCGGAGGGAGAGCGGCCG GATG GAGCGGAGGAT
Pufferfish PEA3 Promoter TGA G - - - - - CAG TTTGA TGGTCA C - - - - - GCA CCGCCG - - - - -

Mouse PEA3 Promoter G A A A G G C G G A T A C T T G G A C C A G C G A G T G C C C T A C A C C T T C T G C A G C G T G A
Pufferfish PEA3 Promoter - - - - - C A G C - - - - - C T T C C G C A G - - - - - A

Mouse PEA3 Promoter G C C C C G C G C G G C C T C C A C C C C G C C C C G C A C C C G C C C C T A
Pufferfish PEA3 Promoter A - - - T G A C A A C A G C T G C C A T T T C G C - - - - - T G

Mouse PEA3 Promoter C T C G C C C G C G A G C C C C C T C C C G G G G T C T A G T G G G T C A G G A C C G C C T G G
Pufferfish PEA3 Promoter A T T G A T C T G C - - - C C C T T T C T C A - - - - -

Mouse PEA3 Promoter C C C C T G A G C C A C C C G G G A C C C C A C C C C A A C C C A A G C T A G A C T C C T C A
Pufferfish PEA3 Promoter C T T C T G - - - - - C T C T C G

Exon 3

Mouse PEA3 Promoter G G G T G A T T C G G G C G A T T C T T C C C C C T C C T C C A G C A G A A A T C T C C C G G A A
Pufferfish PEA3 Promoter G G - - - T T T G T G C - - - - - A G A A A T C G C A T G G A A

Mouse PEA3 Promoter A T G G G A G C T T G G G C G A A G C G C T G A T G G T C C C G C A G G G A A A G C T C A T G G A C
Pufferfish PEA3 Promoter A T G G G C C C C T A A A C A G A C T G T T A A T G G C A G C C A A G A G G A A A T A C A T G G A A

Mouse PEA3 Promoter C C G G G C T C C C T G C C G C C T T C C G A C T C A G A A G G T A A A G G A A A C A G A C A A A C
Pufferfish PEA3 Promoter A C A G A - - - G C T G C C A C C T C A G G A A T C C G A A G G T A A A A - - - - -

Mouse PEA3 Promoter T G C G A G N C C T C T G C T T A A T C C A C G T A C A T C C A T C C A A A G C C T T A T C C T G
Pufferfish PEA3 Promoter - - - - - C T C A G C T C A A - - - - - A T C T C A A T A C T T - - - - -

Mouse PEA3 Promoter G A C T T C C C C A A C C C A C G T T C C C C A A A A G T C A A A C C A C T C A T T C T T C T C C
Pufferfish PEA3 Promoter G A C C T C - - - G G C T C A G G T T - - - - - A T G A T G A A C C - - - - -

Mouse PEA3 Promoter G T C C C T T A T C A C C A G T G G G G A G G C T G T A A C C C G G G C T C G C C A G G T C T C T T
Pufferfish PEA3 Promoter - T C - - - - - G G G A C G T G T C T G A G T C G G A C - - - C C A G - - - C T T

Exon 4

Mouse PEA3 Promoter C C C C A T A T T A T C G C C A C C T C T T C A C C A A T C C A A T A T T C A C C C A T C A G A T C
Pufferfish PEA3 Promoter C T C T A A A C T - - - C T G T C T C C C A C - - - T C C - - - - - C A G A T C

Mouse PEA3 Promoter T C T T C C A G G A C C T C A G T C A C T T C C A A G A G A C G T G G C T C G C A G A A G C T C A G
Pufferfish PEA3 Promoter T T T T C A G G A T T T A A G C C A A C T C C A G G A G A C G T G G C T C A C A G A A G G T G A G

Appendix H Two-way Alignment of the Human and Pufferfish PEA3 Promoters

The two-way alignment of the human and pufferfish PEA3 promoters is illustrated. The position of known exons within each sequence has been depicted by boxes on the sequence with the exon number indicated. The start site of transcription has been shown (+1) in the human sequence. The start site of translation has been indicated in each of the sequences.

Human PEA3 Promoter	T C T C T T G G C C A G A A C T C A C A C T T G T C T C T A A A C T G G T T C C A T C C A G C C C A
Pufferfish PEA3 Promoter	- - - T T T G G - - - G A G C T G A T A - - - - - - - - T A A A - - - - - - - - - - - - - - - - - -
Human PEA3 Promoter	A A G C T G G G A T T T G C T T G G G C T T G T T T C T C C T T C C T T G G G C A G G G T T C C T G
Pufferfish PEA3 Promoter	- - - - - - - - - - - - - - - G G G C T - G C T T T T T T T T T C T T - - - - - - - - - - - - - - - - - -
Human PEA3 Promoter	C T A A G C T C A A T T A A T T T G G T C T T T A T T G T A C C T A T T C A C T A C A C A A G C A T
Pufferfish PEA3 Promoter	- C T A T T T T G T C T G T T - - - - - - - - - - - - - - - - - - T
Human PEA3 Promoter	T T C A T G C C A C C C T C A T A A G C A T G C T A A C A G G C T G A G G A G G A C T T T C A G A G
Pufferfish PEA3 Promoter	T T C A T - - - - - T T T A A A A C A A C A C T G A C G - - - C G A G - - - - - - - - T C A A A G
Human PEA3 Promoter	G C C A A A T G A G C T C T G G A G A G C T C T C C A T A G A T A A G G A T G T C T A T A C C A T G
Pufferfish PEA3 Promoter	T T A A A A C A A A C - - - - - A A A - - - - - - - - - A A A A A A A G A T - - - - - - - - - - - - - - - - - -
Human PEA3 Promoter	A G G C A G G A T G A A A A A C A T C C A C A C C A G A G A T A T A G T C A A A G A A T G T T G T A T
Pufferfish PEA3 Promoter	A G A T A G A G A G A A A A A C A T - G T C G G A G A G T T T T - - - C
Human PEA3 Promoter	A T G G C T A A A G G A A C T G A T T C A G G C T A T C A T G G A A G G C T T C C T G G A G G A G A
Pufferfish PEA3 Promoter	G T G - C C A G G G - - - - - - - - - C A G G T T T T A T G A A - - - - - - - - - T G G A G - - - - -
Human PEA3 Promoter	G G G A C A A T A G A C A T G A G C C T T A T A G A A T A T T A C A G C A C C T A T A G A A A G G G
Pufferfish PEA3 Promoter	- - - - - - - - - - - - - - - - - - - C C C T -
Human PEA3 Promoter	A A G C T T A C T T C C A T T T G C A C A A G C A G A T C T C A A T G G A A T G A G C C C T T A T G
Pufferfish PEA3 Promoter	- - - - T T T C T T T C T C T C T C C A C A A G C A G A - - - - - - - - - T C T C - - - - - - - - - - - - - - - T T T T T G T G
Human PEA3 Promoter	A A C A C T A A A C A G C A A T T G T A T G T T T T A G C C T T A G T T C T G G C C A G A C A G G C
Pufferfish PEA3 Promoter	A A - - - T G A A A A A C A A C T G -
Human PEA3 Promoter	C C A A A G A G C T T A C A T G T G T C A T T T C C T T C T C T G A A A C T C C A C T G C C T C A T
Pufferfish PEA3 Promoter	- - G A A A A G C C - A C A - - - - - - - A T - - - - - - - - - - - A A A A C C C A G C A G C T - - - -
Human PEA3 Promoter	C T G T A A A G T G G C G G G T G T T A G G T T A G A T G A A A T T T G G G A A G C C A C T C C T G C
Pufferfish PEA3 Promoter	- - - - - G A T T G G C - - - T G C C G C G C - - - - - - - A A C C T C G G - - - - - - - - - C C C G C

Human PEA3 Promoter G T A C A G G A G A G T G T G A G G G A T T T G A G G C T G T A A A C C C C A G A G G T G C T G A C
Pufferfish PEA3 Promoter - - - - - - - - - A A A C G C G - - - - - - - - - - - - - - - C T G - A A A G C C G A G C C G T G C - - - -

Human PEA3 Promoter T A G C T G A G T G A C C T C A G G A G A G T T A C C A T C C C T C T C T G A G T T T C A T T T T T
Pufferfish PEA3 Promoter - - - - - - - - - A C C C C - - - - - - - - - C C G C C - - C C C T C C C C G A - - - - - - - - - C C T C

Human PEA3 Promoter A A T A T C T G T A A G A T A C A A A C G A T C C A T T T C G C C G G G T T A T G G G A A C G A A A
Pufferfish PEA3 Promoter A G C A - - - - - - - - - - - - - - - C G G C C C G T C T G A C C C G T T T A C G C - - - A C A A A G

Human PEA3 Promoter T G A G C A C G T G G A A G C G G A G C G C T T C A T A A A T G C T A G T T A T T A T T G T T A C C
Pufferfish PEA3 Promoter T G - - C T C - C G A A A G T C C G G C G C - - - G G A G G T G C G A A - - - - - - - - - - - A C T

Human PEA3 Promoter T C T G G G C A C A C A A G C C C G G T C G C C T C C C T C C T G G G C G C A C A C A G A C A T G T
Pufferfish PEA3 Promoter T C - - A G C - - - - - - - - - G C C C G T C C G T C A C - A C A T G -

Human PEA3 Promoter T C A T A C T C C C A C G T A C A C G C A G A T G C A G A G A C G T G T G C A G A C G C C A G C C G
Pufferfish PEA3 Promoter - - - - - - - - - C G C G C A C C C A C A A A C A A A A A A -

Human PEA3 Promoter G A G G A G C A C G C G G C G C A T C G A C C C A A G A C C G T G G C G A G A C G G T G A A G A A G
Pufferfish PEA3 Promoter - A A A A A - - - - - - - - - G G C A G G T C T G T G T A G G G - - - G T G G - G G G C A G C G - - - - G C

Human PEA3 Promoter C C G C G G G T C C T C G G C T T C T C T C T T T T C T C T C G G T T C T T T A T G A A T G A A
Pufferfish PEA3 Promoter C C G C T G G G T A A T C A - - - - - - - - - C T T C C C T C G G T - - - - - - - - - - - - - - -

Human PEA3 Promoter C T C C A G C C G C G A G T T T T A T T T T T T A T G A A T G G G A G C C G G A G C A C T G A A T
Pufferfish PEA3 Promoter - - - - - - - - - - - - - - - G T T T - - - - - T T T T T T G T G A A T G G A G C T C - - - A T T A - T G C A T

Human PEA3 Promoter G A A G C C G G G A G C C A A C C C C T A C A T T C T G A T T G C C C C G C G G C T G G G A A A
Pufferfish PEA3 Promoter G C A G C C - - - - - - - - - T C C A C G C A G C T C C A T T G A C G C - - - G C A G C T - - - - -

Human PEA3 Promoter C T C C G C C C C C G G C C A G G C C G G A C G A C T C T C C G G A T G - G C G G C G T G C C
Pufferfish PEA3 Promoter - C C C T C C T C T C A T C T G A G T - - - - - - - - - C C C T G C G A G C T G C G C G G A C C A G C T

Human PEA3 Promoter C A G C C G T G C T G G G G T T C C A T T C A C A A A A G T C A A T G A A A C A A A G G G A A A C G
Pufferfish PEA3 Promoter C A G C C C C G A T G A - - - - - - - - - C C A C A A A - - - C G G T T A A - - - - - - - - - - - C G

Exon 2

Human PEA3 Promoter CTG G C A G G C G G A A A C C T C G C C A T C G C C C A G T G A G T C G G A G A A G T T T C G C
Pufferfish PEA3 Promoter C T C G C A G G C G - - - - - C G C - - - - - G T C T - - - - - G T T T T G C

Human PEA3 Promoter C G T C G G A C C C A G A G C C T T G G A C T G C C C C T C C C G C C C A T C G C A G C C
Pufferfish PEA3 Promoter T - - - - - T T T T T C C C C G C C G C G A A C G C G A - -

Exon 2

Human PEA3 Promoter C T C C C T G G A C G G T G T G C G A A C G C A G C C C C C A C C A G G T C T C G G C C
Pufferfish PEA3 Promoter - - - - - A A C G T G - - T G G C T T T C T G A G C A G - - T T T G A - -

In-frame Start Codon

Human PEA3 Promoter C C C G C T T G G G G C C C C G G C C G T G C G G C C G A G G G A G C G G C C G G A T G G A G C G
Pufferfish PEA3 Promoter - - - - - T G G - - - - - T C A - - - C G C A C G G C C G - - - - - C A G C C - - - - - T T C C G C A

In-frame Start Codon

Human PEA3 Promoter G A G G A T G A A A G C C G G A T A C T T G G A C C A G C A A G T G C C C T A C A C C T T C A G C A
Pufferfish PEA3 Promoter G A A - A T G A - - - - - C A A C A G C T G C C - - - - - A T T T T C - - - - -

Human PEA3 Promoter G C G T G A G C G C G C G C G C C G G C C T C C A C G C C C G C C C C G C C C C G C A C C C A G C C C
Pufferfish PEA3 Promoter - - - - - G C T G - A T T G A T C T - - - - - C C C C C G C C C C G C C C C G C A C C C A G C C C

Human PEA3 Promoter C T A C T C T C A C C A C A G C C C C C C T C C C C G C A G T C C C A G C G G A G T C C T C C C C C
Pufferfish PEA3 Promoter C T - T T C T C A C T - - - - - T C T G C T C T C C G - - - - -

Human PEA3 Promoter A N A C T A A G C G C C A C A A G G T G A C T C G C G G G C A T T C T C C C C T C T T C T A G N A G
Pufferfish PEA3 Promoter - - - - - G G T T G T G - - - - - C A G

Exon 3

Human PEA3 Promoter A A A T C G C C C G G A A A T G G G A G C T T G C G C G A A G C C T G A T C G G C C C G C T G G G
Pufferfish PEA3 Promoter A A A T C G C A T G G A A A T G G G C C C C T A A A C A G A C T G T T A A T G G C A G C C A A G A G

Human PEA3 Promoter G A A G C T C A T G G A C C C G G C T C C C T G C C G C C C C T C G A C T C T G A A G G T A A A G A
Pufferfish PEA3 Promoter G A A A T A C A T G G A A A C A G A G - - C T G C C A C C T C A G G A A T C C G A A G G T A A A A -

Human PEA3 Promoter G C A A G A G A G A G A A A C C G C G A C C C T C C T C G C C A C A C A C A C A C A C A C A C
Pufferfish PEA3 Promoter - - - - - C T C A G C T - - - - -

Human PEA3 Promoter A C A C A C A C A C A C A C A C A C A C A C G G C C G A T T C T G G A A C C C C G C C T G T T T C
Pufferfish PEA3 Promoter - - - - - C A - - - - - A A T C T C - - - - -

Human PEA3 Promoter C C C A A A A A A C T C G C C C T C C A A G G G A T T C T G G T C A T T C T T C T C G G T C C C T A
Pufferfish PEA3 Promoter - - - - A A T A C T T G A C C T C - - - - G G C T C A G G T T A T - - - - G A T G - - - A

Human PEA3 Promoter C T C C C A A G G A G G G G A G T C A C C G G G T C T T G C T A G G G C C C C T T C T C T C T A C
Pufferfish PEA3 Promoter A C C T C - - - - G G G A C G T G T C T G A G T C - - - - G G A C C C A G C T T C T C T A A -

Human PEA3 Promoter C G T C C C C T C C A C T C A C C T C C A A T T C C A C T A T T T A C C C C T C A G A T C T C T T
Pufferfish PEA3 Promoter - - - - - - - - A C T C T G T C T C C - - - - C C A C T - - - - C C C A G A T C T T T T

Exon 4

Human PEA3 Promoter C C A G G A T C T A A G T C A C T T C C A G G A G A C G T G G C T C G C T G A A G C T G A G
Pufferfish PEA3 Promoter T C A G G A T T T A A G C C A A C T C C A G G A G A C G T G G C T C A C A G A A G G T G A G

Appendix I Two-way Alignment of the Chicken and Pufferfish PEA3 Promoters

The two-way alignment of the chicken and pufferfish PEA3 promoters is illustrated. The position of known exons within each sequence has been depicted by boxes on the sequence with the exon number indicated. The start site of translation has been indicated in each of the sequences.

Chicken PEA3 Promoter CATTATATCTTTAAACCAGGAAATGCAGCAATTAAATAACCAAGAGTTAAATGG
Pufferfish PEA3 Promoter CATTCCAGCACTTCCGATTGGC-CCGCGTGGGGAG--AGGAG--GGGGGG

Chicken PEA3 Promoter CTATTTTGCAAATGACCTGTATAGCAGAAAGTGCAAAGCAAGCAAGGAGGGGT
Pufferfish PEA3 Promoter TTCATTTCATAA-----GTG-AGCGTA---TTCAAAGGGGACAAAGGGG--

Chicken PEA3 Promoter TCTGGTGTCTGCATAGAGACTTTTGCAGGACACTTAAGTATGAAATTGCTTAAAC
Pufferfish PEA3 Promoter -----AGAAATTTC-----GAGCTCTTGAAGAAGAGGAA--ATCCAC

Chicken PEA3 Promoter AGCAATCTAAACTTGGGGACGTATTCTCTTGCAGCTTCCAGTAG-CATTC
Pufferfish PEA3 Promoter AATCAGGAGGCAGAAACCTT-TATCCCGAATCTGTCCGGCAGGCAGCCG

Chicken PEA3 Promoter GATTGAA GTTAACTGCTTGAACA AATTGCTACAACAATATCTGTATAAT
Pufferfish PEA3 Promoter TGACGGTGC AACT-CTAGAATTATATACGTCCAGAA--ATGCGTAAAAA

Chicken PEA3 Promoter TCTGTATATCTCTGTATAATAATATCCCAAGTAATATCTGGGGATGCGCTATA
Pufferfish PEA3 Promoter GCTGTTCA-----ACC TGT TGA CTGGGGGAGAT-----AAACGCGCTGAA

Chicken PEA3 Promoter AGTTAAGCATGTTCCTTGT TTTGCCCCCATGAACAGATTAACA CCGCTCTTTTG
Pufferfish PEA3 Promoter AGCCGAGCCGTGCACCCCCCGCCCTC-----CCCGACCTCA

Chicken PEA3 Promoter GGCCTGTTTTCTGTGACAGAGTGGTTCTGTGTAATGCTGAAGCTGCTTT
Pufferfish PEA3 Promoter GCACCGGCCCGTCTGAC-----CGCTTACGCAACAAGTGCCTCCG

Chicken PEA3 Promoter TTTTTCACGCAGAAAGATTAATAACTTAACA CTGAATACTTTTCCATC
Pufferfish PEA3 Promoter AAAAGTCCGCGCGGAGGTGCGAA---ACTTCAAGC-----GCCGTCCGTC

Chicken PEA3 Promoter ATACA C GTGTCA TGCTGATCGCA TTGGCTGCAGAGAAAA CATGAC AAAGA
Pufferfish PEA3 Promoter ACA CATGC-----CGCAC---CCACAACA AAAA-----AAAAA

Chicken PEA3 Promoter TGCTCAGTTTCTATA TAGTATT AAGA GTATAATT GTGCTCATCCCTTC CTG
Pufferfish PEA3 Promoter -----AAAGGCAG---GT-----CTG

Chicken PEA3 Promoter TGCAGC CAA GAGTAA AACTGG TTTTGGTTGCTTGGGCTT TAAT TAT TCTAA
Pufferfish PEA3 Promoter TGTAG---GGGTGGGG---GGCAGCGGCCCGCTGGG---TAATCATCTC--

Chicken PEA3 Promoter C T C C A G G T G C A T G A G T G A G A T T T T T T T C C A G A T C T T G T T G G A A C T T G A A
Pufferfish PEA3 Promoter C C C T C G G T G - - - - - T T T T T T T - - - - - T T G T - - G A A T - - - - G

Chicken PEA3 Promoter G A A C T T C C C G G G A A C A G G G C T G A A A A T A T C C T T A A G A C T C T T A G A A A A T G
Pufferfish PEA3 Promoter G A G C T C A T T A T G C A T G C A G C C - - - - - T C C A C G C A G C T C C - - - - A T T G

Chicken PEA3 Promoter A C T G C A G T G T G G G T G A G A A C T G G C A G G C C C G T C G T G T G C G C G C T G G G T
Pufferfish PEA3 Promoter A C - G C - - - - - G C A G C T C C C T C C - - - T C T C A T - C T G A G T

Chicken PEA3 Promoter A G G A G G C T G A A A G T G C A G G C T G A G C G T G C T C G G T C G G C G G C G G T T C G A G C
Pufferfish PEA3 Promoter C - - - - C C - - - - T G C G A G C T G C G C G G A C C A G C T C A G C C C C G A T - - G A C C

Chicken PEA3 Promoter T C C T T T G C A G C C T C T C A T G A A T G A A C T T C C A G A G T T T T T T A T G A A T G G G
Pufferfish PEA3 Promoter A C A - - - - A A C G G T T A A C G A A T G A T T A T A A G A - - - - - C G A T G G A

In-frame Start Codons

Chicken PEA3 Promoter C G C T C C G T C C T T G A A T G A A C A C C A G T G C G C C G C G C T C T G A T T G G C C C G G
Pufferfish PEA3 Promoter T A T T - - - - - T G G A C C A G C A - - A G T G C C T T A C A C T T T A - - - - - G

Chicken PEA3 Promoter C G G A T A A A G C C A C G C C C A C T T T C C C C C A G C G T G C T G C G A T T G G T G G A C G
Pufferfish PEA3 Promoter C A A A T G - - - - T A C G - - T A T - - - - - G G A T G C A G C - - T C G A C C G T C -

Chicken PEA3 Promoter C G C C C A G C C C A G G G C T T C A T T C A C G A A G T C A A T G A A A C A A A G G A G C G G G
Pufferfish PEA3 Promoter C G G C T C G G - - C A C G G C T T A T T T G T T - - - T T A A T - - - - - C G T C

Chicken PEA3 Promoter G A G C C G A G C T C G T G A A G G C G G C G A T G C G C C A A T G G G A G G C C G C C C C G G G
Pufferfish PEA3 Promoter G C G T C G C G C G T C C G A G C G - A G T T T T G C T C C G - T G G C A G - - C A G C C C T G T G

Chicken PEA3 Promoter C G C T C A A C G G C C T G A A C G G C C G C G C G G C G G G G C T G C G T G G G G C T C C G C A
Pufferfish PEA3 Promoter T G T G T G - C G - C G T G T G T G - - T G C G C G C G C G A G A G C G A G T G - - - - - T G C G

Chicken PEA3 Promoter T G G G C C G G G A G T C G C C A C C G G T G A G G C G C G G G C C G A G G G T C G G G C T
Pufferfish PEA3 Promoter T G T G T - - G T G T T T G T - - - - - G T G C G A C G C C G A A - - - - - C T

Chicken PEA3 Promoter C G T C C T G G G A C C G G G G A G G G C T C A T T C T C A G G G T C G G G G A G T T C T G T C T C
Pufferfish PEA3 Promoter C G T C C T A A A A C T A A G T G A A A - - - A T - - - - - G T T G C G - - - - C T A A C G T C

Chicken PEA3 Promoter C T C T C G C T T T T C G T G C A G A C C T C T T C C A G G A C T T G A G C C A G T T C C A G G A G
Pufferfish PEA3 Promoter T C C C A C T C C C C - - - A G A T C T T T T C A G G A T T A A G C C A A C T C C A G G A G

Exon 4

Chicken PEA3 Promoter A C T T G G C T G A C A G A A G G T A G G
Pufferfish PEA3 Promoter A C G T G G C T C A C A G A A G G T G A G

801 0552