Cloning and Characterization of the Promoters of Ryanodine Receptors 1a and 1b (RyR1a/RyR1b) in *Danio rerio*

Rebecca Vanderhooft

Supervisor: Dr. J. Franck

A thesis submitted in partial fulfillment of the Honours Thesis (05.4111/6) Course

Department of Biology

The University of Winnipeg

2010

Abstract

Ryanodine receptors (RyRs) are large homotetrameric proteins, which release calcium from intracellular stores, a critical step in the pathway leading to muscle contraction. In the zebrafish model organism (*Danio rerio*), RyR proteins are encoded by five genes: RyR1a, RyR1b, RyR2a, RyR2b and RyR3. The genes are thought to have arisen from a series of whole and partial genome duplication events. Two duplicate genes can be subfunctionalized, meaning that ancestral expression is divided between the two copies. This is the apparent case with the RyR1 genes of zebrafish in which the RyR1a gene is discretely expressed in slow-twitch muscle and the RyR1b gene is expressed in fast-twitch muscle. Subfunctionalization is caused by changes in *cis*-regulatory elements including the 5' proximal promoter, the adjacent sequence that controls expression via binding of transcription factors. Thus, it is predicted that sequence divergence in the RyR1a and RyR1b promoters has led to the subfunctionalization of these genes. To locate important regulatory elements and test this hypothesis, 10 kb of the RyR1a and RyR1b upstream sequence obtained from databases have been analyzed using comparative genomics tools, as non-coding sequence conservation is typically a marker of regulatory function. Three candidate regulatory regions within the RyR1b upstream sequence have been identified, though there is no obvious conservation within the RyR1a sequence. Furthermore, luciferase vector constructs containing the proximal promoter region of both genes have been created. This work provides the foundation for functional analysis of the RyR1a and RyR1b promoters, raises further questions for exploration and highlights the limitations of sequence conservation as a predictor of regulatory function.

Acknowledgements

Throughout the course of this project, I have had the pleasure of working with many outstanding scientists. First off, I owe a huge thank you to my supervisor, Dr. Jens Franck, who has been incredibly patient and supportive over my tenure in his lab. Thank you also to Dr. Sara Good-Avila and Dr. Ed Byard, for their insight and willingness to edit and discuss my work, and to Dr. Murray Wiegand, for his helpful feedback and organization of the course.

All of my peers in Dr. Franck's lab have provided much appreciated comraderie and advice. Thank you in particular to Siavash Darbandi for maintaining a steady supply of materials required for this project. I have had the great opportunity to collaborate with Ian Kasloff this year on RyR promoters; Ian, thank you for sharing the work and your ideas with me.

My family and friends have been incredibly encouraging and supportive throughout my entire university degree, but particularly in the last year – I cannot thank you enough. Daniel, I am so grateful to have shared this experience with you.

Table of Contents

Abstract ii
Acknowledgements iii
Table of Contents iv
List of Tables v
List of Figures vi
List of Appendices vii
1. Introduction 1
  1.1 Ryanodine Receptor Structure and Function 1
  1.2 Molecular Evolution of Ryanodine Receptors 2
  1.3 Gene Duplication and Molecular Evolution 4
  1.4 Subfunctionalization of RyR1 Paralogues 4
  1.5 Subfunctionalization by *cis*-Regulatory Divergence 6
  1.6 Characteristics of General, Muscle-Specific and RyR Promoters 7
  1.7 Identification of TFBSs 9
  1.8 Significance of RyR1 Genes 9
  1.9 Objectives 10
2. Materials and Methods 13
  2.1 Database Identification of Upstream Sequences: 13
  2.2 Exon-Intron Comparison: 13
  2.3 *In silico* Comparative Genomics: 14
  2.4 Primer Design 15
  2.5 Genomic DNA Extraction: 16
  2.6 PCR Amplification: 17
  2.7 Ligation and Cloning, Purification and Sequencing: 18
  2.8 Construction of RyR1 promoter-pGL4 Constructs/Subcloning into pGL4: 19
3. Results 21
  3.1 Exon-Intron Splicing Comparison 21
  3.2 RyR1a and RyR1b Individual TFBSs 22
  3.3 Sequence Conservation among Teleost RyR1a and RyR1b Promoters 24
  3.4 Conservation of TFBSs among Teleost RyR1a and RyR1b Promoters 27
  3.5 Sequence/TFBS Comparison between Zebrafish RyR1a and RyR1b Promoters 29
  3.6 RyR1a and RyR1b Proximal Promoter Sequence Analysis 30
4. Discussion 34
5. Conclusions 41
6. References 42
7. Appendices 45

List of Tables

1. Primer list 15
2. RyR1a and RyR1b 10 kb promoter MatInspector results. 21
3. Skeletal muscle TFBS families and matrices used for VISTA analysis. 22

List of Figures

1. Molecular evolutionary scheme of the RyR gene family 3
2. RyR1a and RyR1b expression is spatially segregated in fish 5
3. Subfunctionalization occurs via *cis*-regulatory divergence 6
4. Overall experimental design 11
5. PCR detect of RyR1b promoter clones 17
6. Conservation of exon-intron junctions provides support for orthology of the identified zebrafish RyR1a and RyR1b genes 20
7. VISTA analysis: the zebrafish RyR1a promoter and other teleost RyR1a promoters are not conserved.. 24
8. VISTA analysis: the zebrafish RyR1b promoter shows limited conservation with other teleost RyR1b promoters 25
9. Some TFBS are conserved between the zebrafish and takifugu 5' RyR1b sequences 26-27
10. The zebrafish RyR1a and RyR1b promoters are not highly conserved 28
11. Some TFBSs are conserved between the zebrafish RyR1a and RyR1b sequences 29
12. RyR1a proximal promoter as cloned into the pGL4 expression vector 31
13. RyR1b proximal promoter as cloned into the pGL4 expression vector 32

List of Appendices

1. Transcription Factor Information 44
2. Proximal Promoter (2000 bp) Restriction Enzyme Maps 45
2. pGEM-T Easy Vector Maps 46
3. pGL4 Vector Maps 47
4. Ladder (1 kb) 48

Abbreviations & Definitions

RyR:	Ryanodine receptor
SR:	Sarcoplasmic reticulum
DICR:	Depolarization induced calcium release
CICR:	Calcium induced calcium release
WGD:	Whole genome duplication
SGD/PGD:	Single gene duplication/partial genome duplication
PCR:	Polymerase chain reaction
TFBS:	Transcription factor binding site
bp:	Base pair
kb:	Kilobase (1000 bases)
UTR:	Untranslated region
EST:	Expressed sequence tag
HEPES:	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
dNTPs:	Dinucleotide triphosphates
MCS:	Multiple cloning site
ATP:	Adenosine triphosphate
dTT:	Dithiothreitol
SNP:	Single nucleotide polymorphism

Paralogues: related genes within a single species genome, arising from duplication

Orthologues: related genes in separate species genomes, due to common ancestry

Synteny: blocks of genes on a chromosome

EST: short DNA sequence reads obtained by random sequencing of cDNA, used to extend genes to their full length

pGEM-T Easy: a commonly utilised research vector (plasmid), used here for storage and sequencing of PCR products

Contig: DNA sequence constructed from overlapping DNA sequence reads from a single genetic source

pGL4.24: a vector used for promoter analyses, containing an MCS upstream of a firefly luciferase gene with a minimal promoter

Position weight matrix: nucleotide pattern representing a transcription factor binding site; each nucleotide is given a score based on conservation in functionally confirmed TFBSs

Microsatellite: short, repeating segments of DNA (repeating unit = 1-5 bp)

1. Introduction

1.1 Ryanodine Receptor Structure and Function

Ryanodine receptors (RyRs) are large (2.2 million Dalton), homotetrameric, calcium-channel proteins located in the sarcoplasmic reticulum (SR) membrane of striated muscle and various other tissues (Fill and Copello 2002). RyR calcium-channels are critical in excitation-contraction coupling, the process by which a nerve impulse to the muscle membrane is transduced into muscle contraction, since calcium must enter the cytoplasm for contraction to occur. In resting muscle cells, the concentration of cytoplasmic calcium is lower than both extracellular and SR calcium levels. RyRs open in response to an excitatory nerve impulse, allowing SR-stored calcium to enter the cytoplasm. The released calcium binds to troponin, which subsequently allows myosin and actin cross-bridge formation in the sarcomeres (Fill and Copello 2002).

There are two mechanisms by which RyRs are converted to the open channel conformation. In both cases, a nerve impulse depolarizes the cell membrane, spreading down invaginations of the membrane called T-tubules, where dihydropyridine receptors (DHPRs) are located. In the first mechanism, DHPRs are mechanically coupled to RyRs in the adjacent SR membrane and, when the T-tubule depolarizes, conformational change of the DHPRs directly opens the RyR channel (depolarization-induced calcium release, DICR). Alternatively, RyRs can be opened by calcium-induced calcium release (CICR). In this mechanism, DHPRs mediate the movement of a small amount of calcium from extracellular stores to the cytoplasm, where it binds directly to RyRs and induces them to open, allowing the release of the majority of calcium from the SR (Copello *et al*. 2007).

1.2 Molecular Evolution of Ryanodine Receptors

In humans and many other vertebrates, three genes, termed RyR1, RyR2 and RyR3, encode RyR proteins (Takeshima *et al*. 1989; Otsu *et al*. 1990; Zorzato *et al*. 1990). While each channel ultimately releases calcium, they are distinguished by the type of mechanism most frequently used to induce opening – RyR1 via DICR, RyR2 via CICR and RyR3 by DICR – and by their expression pattern. RyR1 is expressed in skeletal muscle, RyR2 is expressed in cardiac muscle and RyR3 is ubiquitously expressed, though at an increased level in brain tissue (Fill and Coppello 2002).

Invertebrates, such as *Caenohabditis elegans* and *Drosophila melanogaster*, and basal vertebrates, such as amphioxus, have only one RyR gene (Sorrentino *et al*. 2000), raising evolutionary questions about this gene family. As indicated in Figure 1, the RyR genes in humans are believed to have originated from a whole genome duplication (WGD) early in the vertebrate lineage, followed by a partial genome duplication (Meyer and van de Peer 2005; Franck *et al*. in preparation), based on information from other gene families, sequence similarity of the RyRs and comparative synteny analysis.

An additional well-documented duplication event has occurred in teleost fish. Evidence from numerous gene families, including Hox genes, predicts a 350 million year old WGD in this group (Meyer and van de Peer 2005). If all genes were maintained, this would implicate a six RyR gene complement in teleost fish. Zebrafish (*Danio rerio*), the most commonly used teleost model organism (Ingham 2009), have five RyR proteins: RyR1a, RyR1b, RyR2a, RyR2b and RyR3 (Hirata *et al*. 2007). The second RyR3 gene may have been lost over evolutionary time.

A recent genomic survey of the bichir (*Polypterus ornatipinni*) genome, a basal ray-finned fish that evolved prior to the WGD 350 million years ago, revealed the presence of four RyR genes, similar to RyR1a, RyR1b, RyR2 and RyR3 in zebrafish (Darbandi and Franck 2009). This result, combined with a lack of conserved synteny between RyR1a and RyR1b paralogues, leads to the predictions that RyR1a and RyR1b evolved as the result of single gene duplication prior to divergence of the ray-finned fishes and the teleost WGD, and that the hypothesized RyR1c and RyR1d genes arising from the WGD were subsequently not maintained (Franck *et al*. in preparation; Figure 1).

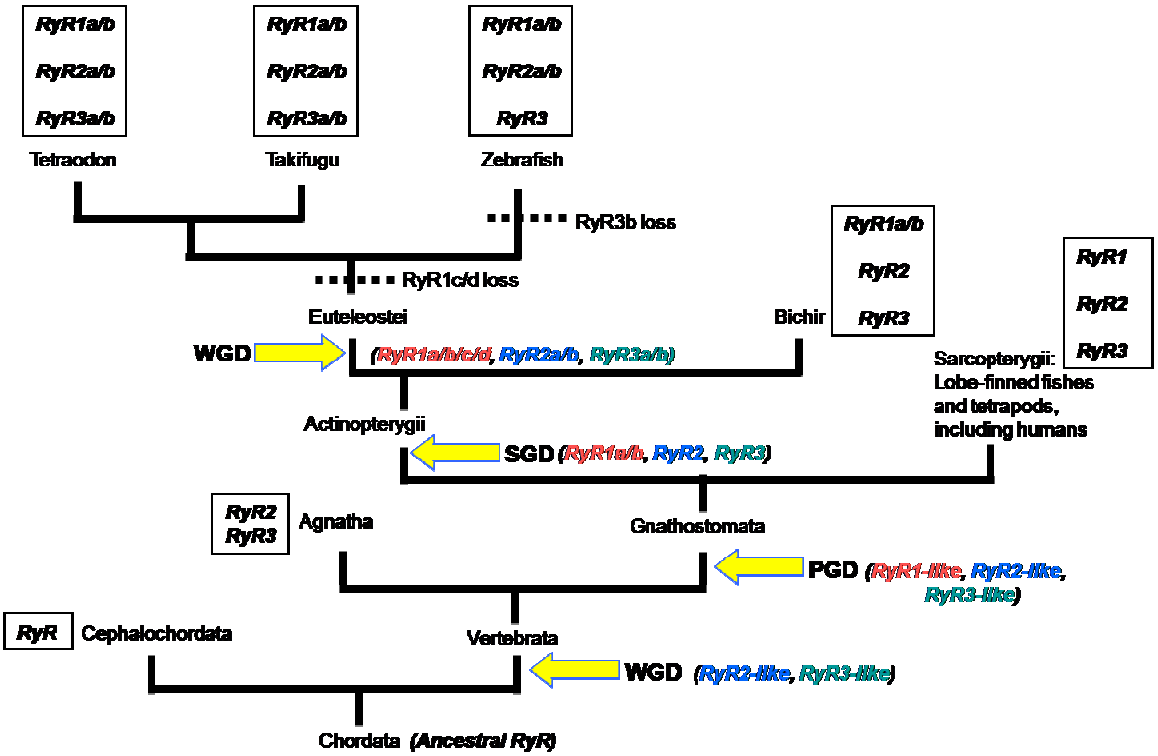


Figure 1. Molecular evolutionary scheme of the RyR gene family. Whole genome duplications (WGD), partial genome (PGD) and single gene duplications (SGD) have established the vertebrate RyR gene complements. Humans fall within Sarcopterygii with three RyRs.

1.3 Gene Duplication and Molecular Evolution

Following gene or genome duplications, the two resulting daughter genes can take one of several routes. While one maintains the ancestral function, the other copy can take on an entirely new function (neofunctionalization) or accumulate deleterious mutations and cease to code for a functional protein (nonfunctionalization); the latter is believed to be the case with the second RyR3 gene in zebrafish and RyR1c and RyR1d among teleosts. In the case of some duplicated genes, the ancestral function is partitioned between the two genes in a process called subfunctionalization (MacCarthy and Bergman 2007). Following a subfunctionalization event, the two paralogues maintain essentially identical functions but are divided in their expression. The two are utilised either at different times (temporal subfunctionalization) or in different tissues (spatial subfunctionalization), whereas the combined expression had been previously accounted for by the single gene (Kleinjan *et al*. 2008).

1.4 Subfunctionalization of RyR1 Paralogues

Overall expression of the RyR genes in teleost fish is similar to that in humans: RyR1 genes are expressed in skeletal muscle, RyR2 genes are expressed in cardiac muscle and RyR3 is ubiquitously expressed (Darbandi and Franck 2009). More specifically, the RyR1a and RyR1b genes are discretely expressed in slow-twitch and fast-twitch skeletal muscle of fish, respectively (Franck *et al*. 1998). Hirata *et al*. (2007) showed that this differential expression of RyR1 genes is established early in development (Figure 2). The fish RyR1a/RyR1b expression pattern is a clear example of spatial subfunctionalization following gene duplication.

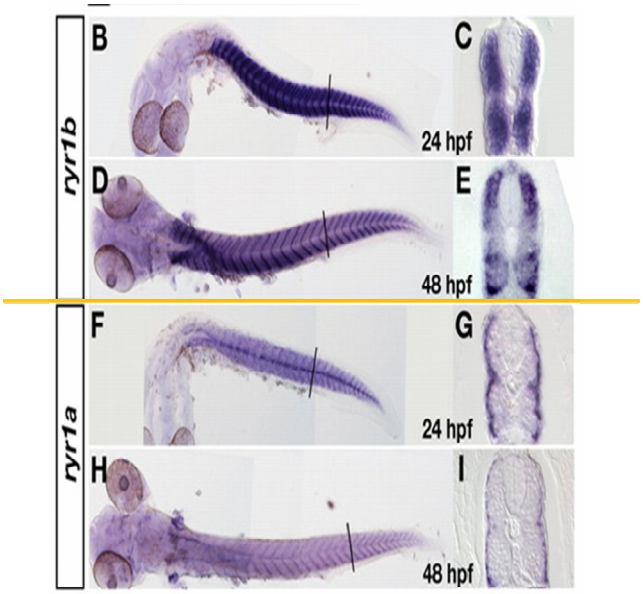


Figure 2. RyR1a and RyR1b expression is spatially segregated in fish. Shown are *in situ* hybridization results in zebrafish embryos using *ryr1b* mRNA probes (B-E) and *ryr1a* mRNA probes (F-I). Hpf = hours post fertilization. Black perpendicular lines in panels B, D, F and H correspond to cross-section in adjacent panels C, E, G and I. RyR1b is expressed in deep, fast-twitch muscles while RyR1a is expressed in shallow, slow-twitch muscles (adapted from Hirata *et al*. 2007).

A recent quantitative real-time polymerase chain reaction (PCR) study showed that the RyR1a and RyR1b orthologues in bichir also share this distinct slow-twitch, fast-twitch expression pattern (Darbandi and Franck 2009). Thus, it appears that subfunctionalization of RyR1a and RyR1b occurred at the base of the ray-finned fish lineage, after their hypothesized SGD and prior to the teleost-specific WGD 350 mya (Meyer and van de Peer 2005; Franck *et al*. in preparation). While it is clear that subfunctionalization has occurred between RyR1a and RyR1b, the underlying molecular mechanisms that maintain this segregation of expression have not been examined.

1.5 Subfunctionalization by *cis*-Regulatory Divergence

Since subfunctionalization results in the division of expression between two paralogues, it is hypothesized that it occurs by sequence divergence of *cis*-regulatory elements – regions responsible for controlling transcription – of the daughter genes (MacCarthy and Bergman 2007). Following gene duplication, the two copies presumably share regulatory elements with redundant transcription and activity of the protein products. Subsequently, regulatory mutations that lead to partial loss of expression of one copy are tolerated, since the second copy will still be expressed and perform the function in the tissue or at the time required (Figure 3). Over time, complementary mutations within common regulatory elements lead to partitioning of the ancestral expression and function between two duplicate genes (MacCarthy and Bergman 2007).

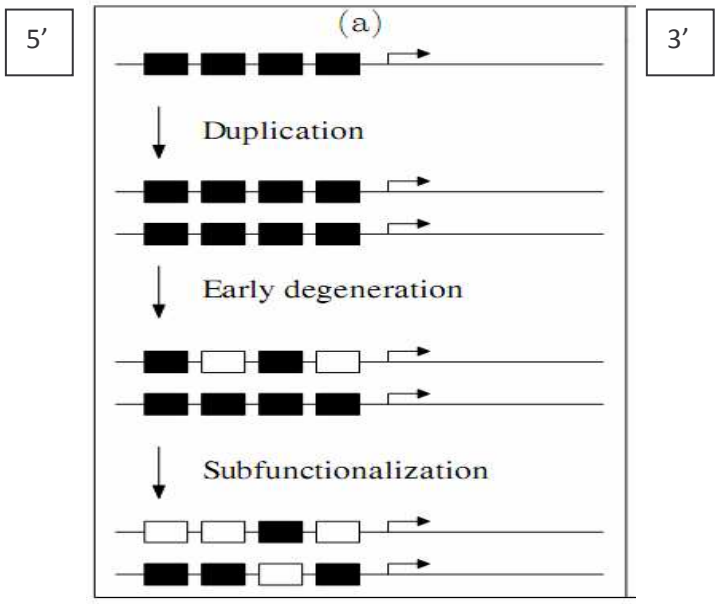


Figure 3. Subfunctionalization occurs via *cis*-regulatory divergence. Black boxes indicate functional *cis*-regulatory elements upstream of a gene (transcription represented by arrow). White boxes indicate mutations leading to loss of that regulatory element. Over time, complementary mutations divide the ancestral expression between two genes (adapted from MacCarthy and Bergman 2007).

This process has been termed *cis*-regulatory divergence as it results in a division of the ancestral *cis*-regulatory elements between the two daughter genes. Thus, regulatory elements are not expected to be identical between subfunctionalized genes; rather, some ancestral elements should be present for one paralogue, while other ancestral elements are maintained near the other paralogue. This does not imply that all regulatory elements need be divided (Kleinjan *et al*. 2008); in fact, conservation of some elements is still expected as, for example, RyR1a and RyR1b are still both expressed in skeletal muscle and require elements that drive expression in that tissue.

Subfunctionalization of *pax6a* and *pax6b* genes by *cis*-regulatory divergence has recently been characterized at the sequence level in zebrafish. In this case, some highly conserved ancestral *pax6* mammalian elements were found to be present near both *pax6a* and *pax6b*, while others were found near only one paralogue (Kleinjan *et al.* 2008). Subfunctionalization by *cis*-regulatory divergence continues to be a topic of intense research across many groups of organisms. I hypothesize that *cis*-regulatory divergence is also responsible for the differential expression of the fish RyR1a and RyR1b genes.

1.6 Characteristics of General, Muscle-Specific and RyR Promoters

Cis-regulatory elements constitute functional clusters of transcription factor binding sites (TFBSs) to which transcription factor proteins bind to initiate transcription (Narlikar and Ovcharenko 2009). These elements can include promoters, enhancers and repressors and can be located upstream, downstream or within genes (Jegga *et al*. 2002; Kleinjan *et al*. 2008). Promoters, specifically, are sequences immediately upstream (5') of genes. One common promoter TFBS is the TATA box, which is present in approximately

25% of human and yeast promoters (Yang *et al*. 2006). The relatively short region immediately upstream of a gene containing the TATA box, if present, and other TFBSs is typically referred to as the proximal promoter, while further upstream is considered distal promoter sequence. In this study, the 2000 base pairs (bp) upstream of RyR1a and RyR1b are defined as proximal.

Muscle-specific promoters, relevant for RyR1a and RyR1b, have been shown to contain binding sites for TFs in the MYOD gene family and an A/T-rich region recognized by the TF MEF2 (see Appendix 1 for full names of all TFs described in this thesis). MEF3 binding sites have also been identified in muscle-specific promoters or enhancers of muscle-specific murine aldolase A, myogenin and cardiac troponin C genes (Dunwoodie *et al*. 1994; Schmoelzl *et al*. 1996). While muscle-specific promoters are typically short, some regulatory elements have been found at greater distances upstream, in the first intron of the gene itself or downstream of the gene (Kleinjan *et al*. 2008).

Two RyR promoters have been experimentally examined. Schmoelzl *et al*. (1996) found that, in the porcine RyR1 promoter, a 552 bp sequence containing a GC-rich Sp1 binding site, an antisense MEF2 binding site and two novel binding sites was sufficient to drive non-tissue specific transcription, even without a TATA box. However, E-box binding sites contained within the first intron were necessary to drive muscle-specific transcription. The rabbit RyR2 promoter has also been characterized and shown to lack a TATA box, while containing an Sp1 binding site and novel binding sites (Ntshida *et al*. 1996). Notably, in both cases, a limited length of the promoter (<1000 bp) was examined. Neither zebrafish nor human RyR promoters have been analyzed to date.

1.7 Identification of TFBSs

In beginning an analysis of promoters, the identification of potential transcription factor binding sites (TFBSs) is essential. Programs which scan sequences for individual TFBSs, however, typically result in outputs with extremely high numbers of putative binding sites, many of which are likely false positives (Jegga *et al*. 2002). To minimize false positives for functional analysis it is recognized that, though non-coding sequence is not typically conserved due to lack of selective constraint, important functional regulatory elements are likely conserved between related but divergent species (Bird *et al*. 2006; Fisher *et al*. 2006; Kleinjan *et al*. 2008). Searching for these 'phylogenetic footprint' sequences, conserved over long periods of evolutionary time – for example, between teleosts or between mammals and teleosts – has become a commonly used strategy for identifying clusters of important TFBSs (Jegga *et al*. 2002). A greater distance is better; for example, human to takifugu (>300 million years) is often used (Fisher *et al*. 2006).

1.8 Significance of RyR1 Genes

Aside from the interest in subfunctionalization, an analysis of RyR1 transcription could have implications for health research. Missense mutations in the RyR1 gene have been implicated in malignant hyperthermia and central core disease (Shmoelzl *et al*. 1996; McCarthy *et al*. 2000). Malignant hyperthermia is an episodic disorder triggered in individuals in response to halogenated anaesthetics and muscle relaxants during anaesthesia. An incident is characterized by hypermetabolism, muscle rigidity and increased temperature among many other symptoms (Fill and Copello 2002). Central core

disease, comparatively, is a congenital myopathy characterized by muscle weakness in infants; it is diagnosed by the presence of few mitochondria in skeletal muscle (McCarthy *et al*. 2000).

In addition, an insertion in an intron of the zebrafish RyR1b gene has been shown to generate aberrant splicing, leading to inefficient muscle contraction with a slow swimming phenotype (Hirata *et al*. 2007). As RyRs play a key role in muscle contraction, there is a good possibility that unidentified mutations or further, aberrant gene expression could be associated with other muscular disorders or degenerative diseases. It is common for mutations in regulatory elements to have an effect on human health (Narlikar and Ovcharenko 2009); thus, it is important to understand RyR1 gene expression. Analyzing RyR1a and RyR1b expression in zebrafish, an organism used to study many other diseases (Ingham 2009), is a good starting point.

1.9 Objectives

This work establishes a basis for complete analysis of transcriptional regulation of the critically important skeletal muscle RyR1 genes, using zebrafish as a model organism. Here, I have conducted an *in silico* analysis of the 10 000 bp (10 kilobase, kb) upstream of the RyR1a and RyR1b genes, using the web-based MatInspector and VISTA applications to scan for TFBSs and candidate regulatory regions using non-coding sequence conservation as a regulatory marker, respectively. I have also created luciferase expression vector promoter constructs containing the proximal 5' 2000 bp promoters of the RyR1a and RyR1b genes, a logical starting point for functional analysis. These constructs will later be tested for their ability to drive transcription via transient

expression analysis (as explored in the ensuing discussion) and if capable, whether transcription is muscle-specific. The *in silico* information presented here, along with results from later expression assays using the constructs I have created, will be utilised to direct creation of additional reporter constructs and carry out deletion analysis of the current constructs in future experiments (Figure 4). Ultimately, important regulatory regions for each of the RyR1a and RyR1b promoters will be identified.

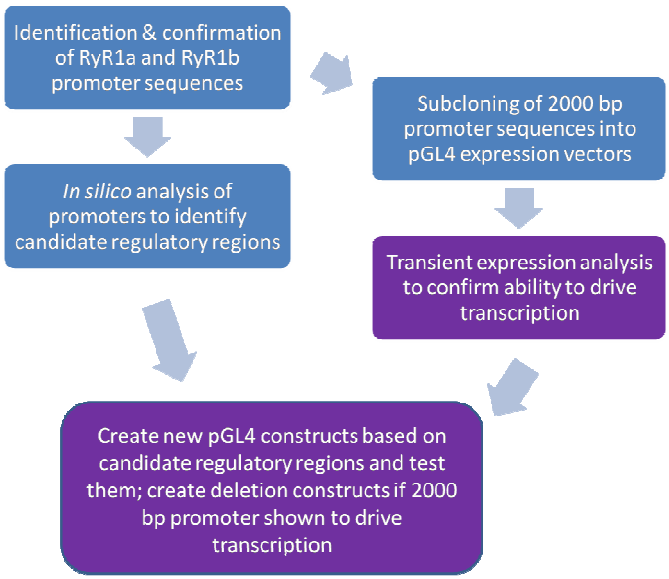


Figure 4. Overall experimental design. Blue boxes were completed in this study. Purple boxes will be addressed in future experiments.

Another goal is to identify evidence of *cis*-regulatory divergence, but at this point, I am limited by incomplete knowledge of ancestral RyR1 promoters. The best evidence for *cis*-regulatory divergence, based solely on comparative sequence analysis, would be sequence divergence between RyR1a and RyR1b, as well as complementary conservation of sequence and predicted TFBSs between zebrafish RyR1a/RyR1b and ancestral RyR1 promoter sequences. I have begun this analysis by comparing the zebrafish RyR1a and RyR1b promoters to selected vertebrate species, including human.

2. Materials and Methods

2.1 Database Identification of Upstream Sequences:

The 5' segment of the RyR1b gene was identified in the latest zebrafish genome assembly (Zv8) in the Ensembl database by querying the RyR1b cDNA transcript cloned by Hirata *et al*. (2007). A gene on chromosome 18 containing the 5' region was identified and confirmed through correspondence with Ensembl staff (ENSDARG00000075679). The 10 kb sequence upstream of the predicted transcription start site (including a 160 bp 5' untranslated region, UTR) was extracted from the database. The RyR1a gene on chromosome 10 is only partially annotated in Zv8, though it was entire in the previous Zv7 assembly, and the 5' portion of the gene is missing. Based on my correspondence with staff at Ensembl, the 5' region is located in the clone BX088525, which has now been manually annotated and will be included in upcoming database releases. I received the 10 kb sequence upstream of the predicted transcription start site (including a 175 bp 5' UTR) from Ensembl staff. In both cases, the 5' UTR was identified by expressed sequence tags (ESTs).

2.2 Exon-Intron Comparison:

The first ten exon-intron boundaries of RyR1 genes in human and several teleost fish were compared to add support for the orthology of the zebrafish RyR1a and RyR1b genes and promoters, since identification of the database sequences was not straight-forward. Exon-intron junctions are typically conserved over evolutionary time and human RyR1 exon-intron splice sites have been experimentally determined (Phillips *et al.* 1996). Queries of the takifugu and tetraodon (both teleost fish) database genome assemblies for

"RyR1" yielded two genes for each species: ENSTRUG00000016954, ENSTRUG00000015260 and ENSTNIG00000019014, ENSTNIG00000010011. Within the database, these genes are called FuguRyR1-1, FuguRyR1-2, OdonRyR1-1 and OdonRyR1-2 respectively, rather than RyR1a and RyR1b. The exon-intron amino acid boundaries identified in Ensembl for the first transcript of these genes, as well as for RyR1b, were manually compared to human RyR1 splice junctions. Zebrafish RyR1a exon sequences were individually translated to determine amino acid splice sites for comparison. In all cases, analysis of the first exon began at the codon for the first methionine amino acid, as transcription start sites are unknown for takifugu and tetraodon. This analysis was also used to identify which of the takifugu and tetraodon RyR1 genes are orthologous to zebrafish RyR1a and RyR1b respectively.

2.3 *In silico* Comparative Genomics:

MatInspector is an online comparative genomics tool which identifies TFBSs using TRANSFAC position weight matrices (Cartharius *et al*. 2005). MatInspector was used to scan the database 10 kb + 5' UTR RyR1a and RyR1b sequences to identify putative transcription factor binding sites; changes identified in the proximal promoter by sequencing (as described in the following methods) were incorporated before scanning. General core promoter element and vertebrate matrices were selected, initially with settings of 75% core similarity and optimized matrix similarity. Results with 90% core similarity and optimized matrix similarity and results with 90% core similarity and optimized + 0.1 matrix similarity were also obtained for comparison. Transcription factor families that were identified in these results and that are identified by MatInspector to be

expressed in skeletal muscle were selected for further analysis. Within each family, the individual TFBS matrices available in VISTA, the next program to be used, were noted (35 total). Two additional transcription factors were selected for VISTA analysis: Sp1, based on its known importance in the porcine and rabbit RyR promoters, and Muscle.

RyR1a and RyR1b promoter sequences were next aligned to takifugu, tetraodon and medaka RyR1a-like or RyR1b-like promoters respectively, using a 100 bp window, minimum 70% similarity and the LAGAN alignment program in VISTA. VISTA is a comparative genomics program that aligns non-coding sequences and identifies conserved TFBSs, in mVISTA and rVISTA respectively (Frazer *et al*. 2004). Comparison to the human RyR1 upstream sequence was included in these two VISTA plots. The alignments were then scanned for the 35 MatInspector-identified skeletal muscle TFBSs using a standard 24 bp sliding window to identify conserved sites for the RyR1a and RyR1b orthologues respectively.

The RyR1a and RyR1b sequences were similarly aligned and scanned in VISTA to identify any conserved regions and TFBSs as part of a preliminary analysis of *cis*-regulatory divergence.

2.4 Primer Design

To confirm and isolate the RyR1a and RyR1b proximal promoter sequences for expression arrays, the Primer3 program was used to design primers targeting the 2000 bp sequence immediately upstream of the RyR1a and RyR1b genes, including the 5' UTR (Rozen and Skaletsky 2000). A 2000 bp subset of the 10 kb sequences was chosen to begin expression analysis since it is practical and there is precedent for the ability of

relatively short muscle-specific promoters to drive transcription (Dunwoodie *et al*. 1994; Schmoelzl *et al*. 1996). Primers were designed to amplify ~500 bp overlapping fragments, as direct sequencing is limited to this length, and the full-length ~2000 bp promoter sequences for cloning (Table 1).

Table 1. Primer list. All primers successfully utilised throughout this study, including those used to sequence pGEM-T and pGL4 plasmids. Primers in bold were used to amplify the full-length 2000 bp proximal promoters. Underlined sequences in the restriction enzyme primers indicate the incorporated restriction enzyme site and added terminal CGC.

Sequence Target	Primers: 5' – 3'			
RyR1a	FA1	TGGGGTCTTAAAATTCATGG	RA2	ACCGCCAACTATTCTCTTC
	FA3	TTGGCCTTTGTGTATGAGTG	RA3	TAACAGGCAGCTCCACTTTC
	FA4	GATGCCAGCACTCATCATTAC	RA4	ATCTCCTCTTCTCCCTCAGC
	FA6	AAATGTTTGCCAAGCTTACG	RA6	TATGCAGGAATCAGCAAGTG
	FA7	**AATTAGGCAACGCGTTATTG**	RA7	AAATGCTAATCCAGCAGGTG
			RA9	**TTCTGCCATGGTACTGGAG**
RyR1b	FB1	TGGTGGCTTGTATGAGTGTG	RB1	TGCTGAAACATGACTTTTGG
	FB2	GTCATGTTTCAGCAGTTTGC	RB2	AGCCGGTTATATCGAGTGC
	FB3	GATATAACCGGCTGCGTTC	RB3	TCTCAAATCGTGTGGATGG
	FB4	TATCTGCTTGCGTGTGTCTC	**RB4**	**CCTTCTCCGAGGGTTATTTG**
	FB5	TGGTGGACTGTGAGTGTCTG	RB5	CACAAACGACAGAAGTACAGGATG
	FB7	CAGACAGACTTTGTTTACTTGC	RB7	TTGGTTTGTCAGAACTAGGG
	FB8	CCATCCACACGATTTGAGAG	RB8	CAACTTGAAGCCCACTCCTC
	FB9	TGGGCGCTATAAACGAATAC	RB9	TGCACAGACACTCACAGTCC
	FB10	AAAACATACCCCTGGAGGTC	RB10	CATCGGAGAGTGGAGTAACG
	FB12	**TGTACTGTTACATGTTGGTGTCC**		
pGEM-T	M13F	CACGACGTTGTAAAACGAC	M13R	GGATAACAATTTCACACAGG
RyR1a + Restriction Enzyme	FA7-KpnI	CGCGGTACC AATTAGGCAACGCGTTATTG	RA9-XhoI	CGCCTCGAG TTCTGCCATGGTACTGGAG
RyR1b + Restriction Enzyme	FB12-KpnI	CGCGGTACC TGTACTGTTACATGTTGGTGTCC	RB4-XhoI	CGCCTCGAG CCTTCTCCGAGGGTTATTTG
pGL4	RV3	CTAGGCAAAATAGGCTGTCCC		

2.5 Genomic DNA Extraction:

Zebrafish were purchased from a local fish store and sacrificed by a protocol approved by the Senate Animal Care Committee issued to Dr. Jens Franck (April 2009). The fish were euthanized in buffered MS222, immediately frozen in liquid nitrogen and

stored at -80°C. For extractions, fish were homogenized and left overnight at 37°C in 1 ml DNAzol reagent supplemented with 100 µg/ml proteinase K. The homogenate was centrifuged for 10 minutes at 10 000 x g, followed by DNA precipitation in 95% ethanol. DNA was washed with 75% ethanol and dissolved in 8 mM sodium hydroxide, followed by neutralization to a pH of 8.0 with 0.1 M HEPES. Several genomic extractions were used throughout these procedures.

2.6 PCR Amplification:

Short ~500 bp fragments and the 5' proximal 2000 bp promoter sequences were amplified and isolated using PCR. PCR reactions were mixed to a final volume of 25 or 50 µl, with each reaction containing 1X PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 1 µM of each primer, 2-2.5 units (U) Taq DNA polymerase (all final concentrations, all supplied by Invitrogen Life Technologies), RNase free water and 1-5 µl genomic DNA template ranging in concentration from 45-145 ng/µl. A reaction lacking DNA template was set up alongside each series as a negative control. All PCR reactions were initially denatured for 9 minutes at 95°C followed by 35 cycles as follows: 1 minute denaturation at 95°C, 1 minute 30 seconds annealing at 53-58°C (depending on the primers), 1 minute extension at 72°C, and a final 7 minute elongation step at 72°C. Products were visualized with ethidium bromide staining under UV light following electrophoresis at 95 Volts for 45-60 minutes in 1% agarose gels. Successfully amplified and correctly sized products were purified from 1% agarose low melting point gels, following electrophoresis at 95 V for 1 hour, using the S.N.A.P. Gel Purification Kit (Invitrogen).

2.7 Ligation and Cloning, Purification and Sequencing:

Some PCR products, including the 2000 bp promoter sequences, were ligated into pGEM-T Easy Vectors (Promega; Appendix 2), by incubating the product at 4°C with 50 ng vector, 3 U ligase and 1X ligase buffer. The ligation reactions were transformed into competent *Escherichia coli* according to manufacturer's protocol (strain JM109, Promega) and grown on LB media plates containing 0.05 mg/ml ampicillin. Plates were incubated at 37°C for 24 hours.

Colonies were randomly selected and checked for successful ligation by PCR. Each selected colony was used to inoculate a 15 µl PCR mix containing 1X PCR buffer, 3 mM $MgCl_2$, 0.25 mM dNTPs, 0.5 µM each of M13F and M13R primers (Table 1), RNase free water and 2 units Taq DNA polymerase (Invitrogen). The PCR cycle was as follows: 5 minutes at 95°C; (34X) 1 minute at 95°C, 30 seconds at 55°C, 1 minute at 72°C. Products were sized on 1% agarose gels (Figure 5) and colonies corresponding to correct PCR products were grown at 37°C in LB broth with 0.05 mg/ml ampicillin.

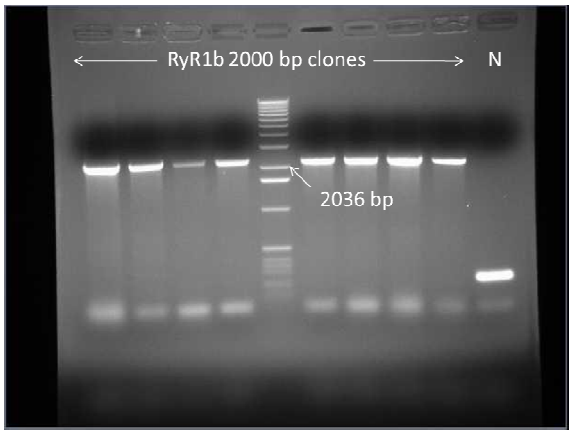


Figure 5. PCR detect of RyR1b promoter clones. M13F and M13R primers amplified ~2000 bp RyR1b sequence + some adjacent pGEM-T vector (about 2300 bp total). N = 300 bp empty plasmid amplified by M13F/M13R. 1% agarose gel stained with ethidium bromide and visualized under ultraviolet light.

Plasmids were purified from overnight bacterial cultures using the PureLink Quick Plasmid Miniprep Kit (Invitrogen), and sent to The Centre for Applied Genomics (TCAG) in Toronto for sequencing on both strands. Some PCR products were direct sequenced without ligation into pGEM-T vectors.

Sequences were edited using SeqMan (Lasergene, DNASTAR), and assembled into a contig spanning each respective promoter region. Contigs were aligned to the database sequence using ClustalX to confirm the correct segment before subcloning.

2.8 Construction of RyR1 Promoter-pGL4 Constructs/Subcloning into pGL4:

Restriction enzyme maps were generated for the sequenced 2000 bp promoters using NEBCutter (Vincze *et al.* 2003; Appendix 3). Based on these maps, KpnI and XhoI restriction enzyme sites were selected in the pGL4.24 firefly luciferase expression vector multiple cloning site (MCS) to insert the promoters and create reporter plasmids (Appendix 4). Since KpnI and XhoI sites are not present in the pGEM-T vector MCS, primers were designed to amplify the promoters out of pGEM-T that incorporated the KpnI and XhoI recognition sites and a terminal CGC. PCR was carried out on 1 ng pGEM-T templates using these primers with subsequent gel-purification, following a protocol identical to section 2.6. The annealing temperature for PCR was 58°C.

The proximal promoters (~100 ng) were digested with 20 ng pGL4.24 vector in a single 25 µl reaction containing 1x Buffer A (Roche Diagnostics), 10 mg/ml bovine serum albumin, 5 U KpnI, and 5 U XhoI. Digests were incubated at 37°C for two hours followed by a 15 minute heat-inactivation period at 65°C.

Following incubation, 4.5 U ligase, ATP and dTT were added directly to digests, the latter at final concentrations of 1 mM and 10 mM respectively. The ligations were incubated overnight at 4°C before transformation, again into *E. coli* strain JM109. Uncut plasmid (20 ng) was transformed alongside the 10 µl ligation mixes as a positive control. Isolated colonies were tested by colony PCR with the terminal promoter primers (FA7/RA9, FB12/RB4 in Table 1), grown in culture and harvested for plasmids as indicated in section 2.7.

To confirm that the isolated pGL4 plasmids included correctly inserted RyR1a and RyR1b promoters, the reporter constructs were sequenced at TCAG. The RV3 primer just upstream of the MCS in pGL4 (Appendix 4, Table 1), the reverse RA9/RB4 terminal primers and 2-3 internal primers per promoter were utilised for sequencing. A full promoter sequence was obtained for each and compared to Ensembl using ClustalX (Larkin *et al*. 2007).

3. Results

3.1 Exon-Intron Splicing Comparison

The alignment used to compare exon-intron boundaries (Figure 6) revealed that the zebrafish RyR1a gene is orthologous to takifugu RyR1-2 and tetraodon RyR1-2, while zebrafish RyR1b is orthologous to takifugu RyR1-1 and tetraodon RyR1-1. These orthologue groups were used in the VISTA analysis.

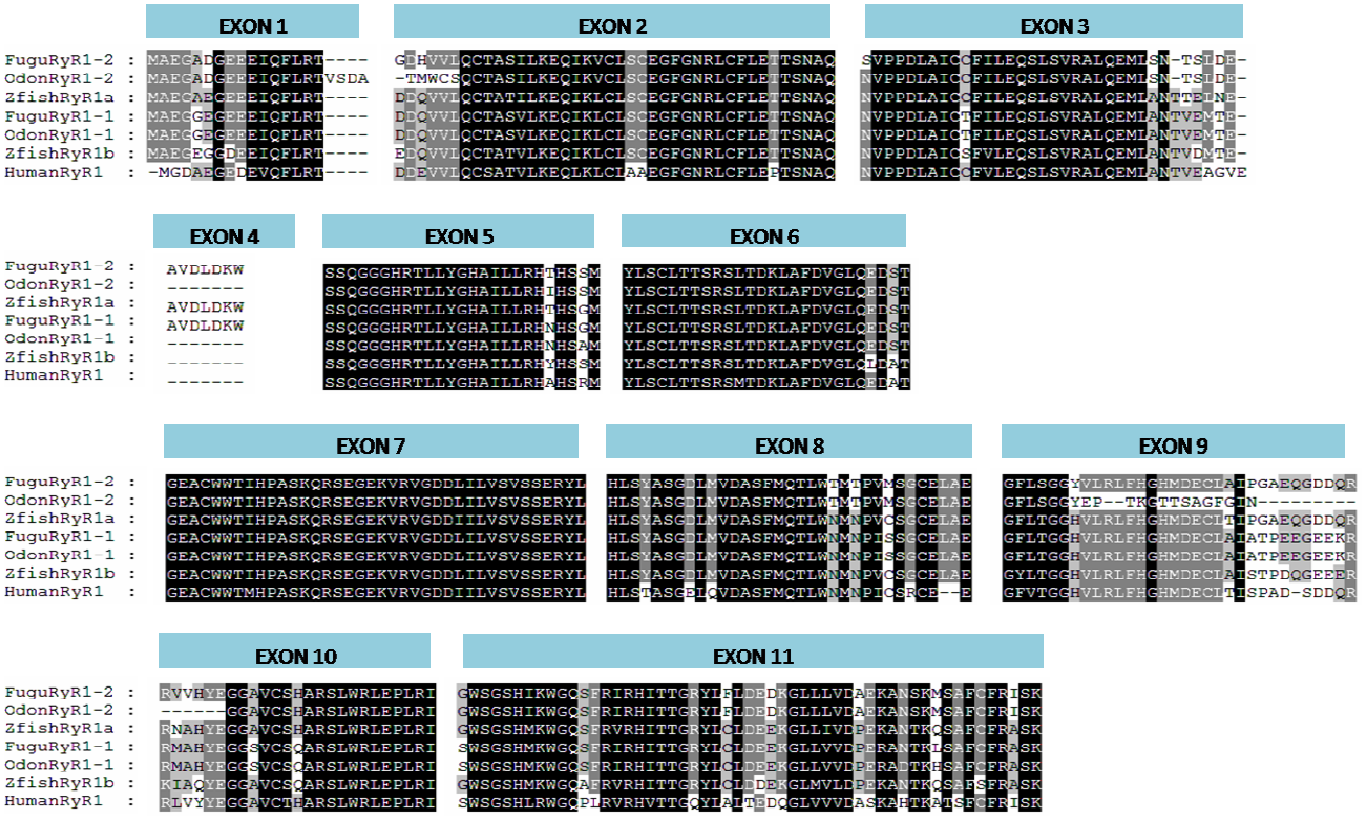


Figure 6. Conservation of exon-intron junctions provides support for orthology of the identified zebrafish RyR1a and RyR1b genes. Splicing pattern for the first eleven exons of teleost RyR1 genes, as indicated in Ensembl for the first transcript of FuguRyR1-2, OdonRyR1-2, FuguRyR1-1, OdonRyR1-1 and ZebrafishRyR1b genes. RyR1a exons were manually translated based on the annotated sequence received from Ensembl staff. Human RyR1 exons have been experimentally determined (Phillips *et al*. 1996). Degree of shading indicates level of conservation per amino acid site.

As observed in Figure 6, the splice sites for the first ten exons are predominantly conserved between takifugu, tetraodon, human and the identified zebrafish RyR1 genes. The only variations are that the tetraodon RyR1a-like sequence varies at the exon 1-2 and exon 9-10 junctions, and both takifugu genes and the zebrafish RyR1b gene have an additional exon. Since these changes do not indicate unusual splicing in the zebrafish genes, exon-intron analysis provides support for the orthology of the identified RyR1a and RyR1b genes and thus, orthology of the identified upstream sequences.

3.2 RyR1a and RyR1b Individual TFBSs

Scanning the RyR1a and RyR1b 10 kb upstream sequence plus the predicted 5' UTRs with MatInspector standard and modified settings yielded the results in Table 2. The high number of TFBS hits is not unexpected, although it complicates further analysis. Scanning with altered stringency levels was performed for comparison; setting higher stringency levels does reduce the number of hits by decreasing false positives, but also introduces an increased frequency of false negatives (Carrhenius *et al*. 2005).

Table 2. **RyR1a and RyR1b 10 kb promoter MatInspector results**. Individual TFBS hits for 10 kb promoters + 5' UTRs with altered settings. The sequences were scanned for vertebrate and general core promoter sites. Core similarity is the required conservation for the most highly conserved part of the binding site (typically 4 bp). Matrix similarity is the required conservation for the entire binding site (typically 10 bp).

MatInspector Settings	TFBS Hits		Skeletal Muscle TFBS Hits	
	RyR1a	RyR1b	RyR1a	RyR1b
0.75 Core Similarity; Optimized Matrix Similarity	3245	3288	312	279
0.90 Core Similarity; Optimized Matrix Similarity	2552	2608	232	228
0.90 Core Similarity; Optimized + 0.1 Matrix Similarity	310	351	33	26

To regulate expression of the skeletal muscle RyR1a and RyR1b genes, transcription factors must be present in skeletal muscle. To simplify further analyses, and since it makes biological sense, attention was focussed on transcription factor families identified by MatInspector to be expressed in skeletal muscle. The numbers indicated in Table 2 do include binding sites for muscle-expressed transcription factors: Table 3 lists these TFBS families and individual matrices in each family available in VISTA, the next program used (see Appendix 1).

Table 3. Skeletal muscle TFBS families and matrices used for VISTA analysis.
Matrices as extracted from the MatInspector database; see Abbreviations and Appendix.

TFBS Family	Position Weight Matrices	
Fork head domain factors (FKHD)	FREAC2	HFH8
	FREAC3	HNF3
	FREAC4	XFD1
	FREAC7	XFD2
	HFH1	XFD3
	HFH3	
Cart-1 cartilage homeoprotein 1 (CART)	ALX4	S8
	CART1	
Pax homeodomain binding sites (PAXH)	PAX4	
	PAX6	
TEA/ATTS DNA binding domain factors (TEAF)	TEF1	
Human and murine ETS1 factors (ETSF)	CETS1P54	GABP
	ELK1	NRF2
	ETS1	PEA3
	ETS2	PU1
Serum response element binding factors (SRFF)	SRF	
Myocyte-specific enhancer binding factors (MEF2)	MEF2	
	RSRFC4	
MEF binding sites (MEF3)	MEF3	
Myoblast determining factors (MYOD)	E47	MYOGENIN
	MYOD	
Pax-3 binding sites (PAX3)	PAX3	
Sp1	GC	
	SP1	

3.3 Sequence Conservation among Teleost RyR1a and RyR1b Promoters

To narrow down putative TFBS sites for functional testing, non-coding sequence conservation was used as an indicator of regulatory function. A series of computational tools in VISTA were employed to identify sequence conservation and to determine which of the muscle-specific MatInspector hits are conserved and clustered. First, alignments of the RyR1a and RyR1b 5' sequences with the 5' sequences of respective orthologous RyR1 genes reveal that the zebrafish 10 kb RyR1a sequence is not conserved with other teleost RyR1a or the human RyR1 sequences. The RyR1b promoter sequences are minimally conserved as well, though there is one site that exceeds 70% similarity over a 100 bp aligned segment and two sites that exceed 70% over segments shorter than the required 100 bp, between zebrafish and all three other teleost fish (Figure 7 and 8). These three putative regulatory regions are located from -9400 to -9100, -5900 to -5300 and within the 5' UTR, with respect to +1 being the RyR1b transcription initiation start site. Conservation of these three regions does not extend to the human RyR1 upstream sequence.

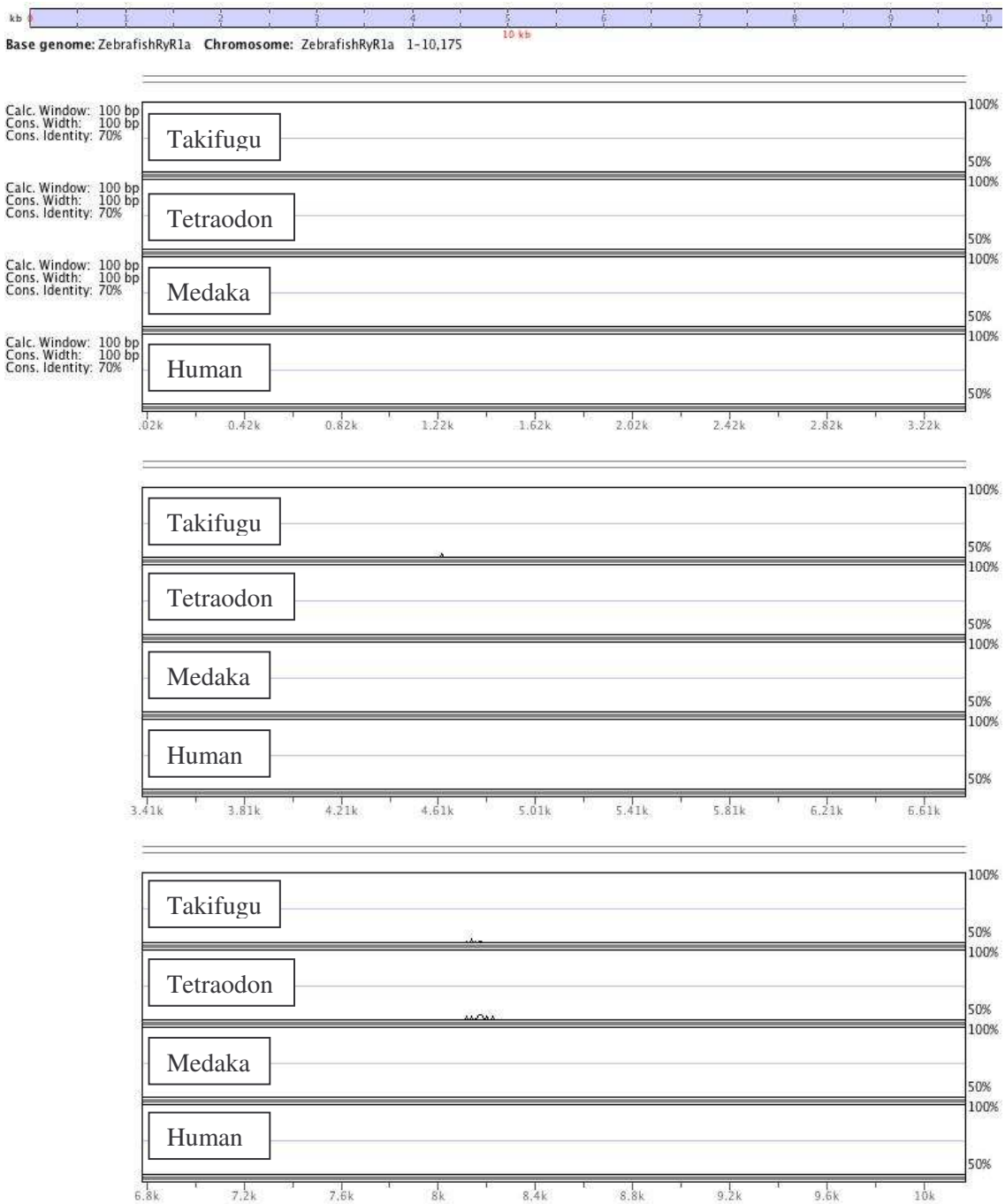


Figure 7. VISTA analysis: the zebrafish RyR1a promoter and other teleost RyR1a promoters are not conserved. Base sequence = zebrafish RyR1a 5' region. Base 0 corresponds to 10 kb upstream of the 5' UTR, base 10 175 is immediately adjacent to first exon. From top: alignment with 5' sequences of takifugu, tetraodon, medaka RyR1a-like genes and human RyR1. Alignment based on sliding 100 bp window; peaks indicate regions conserved above 50% over 100 bp.

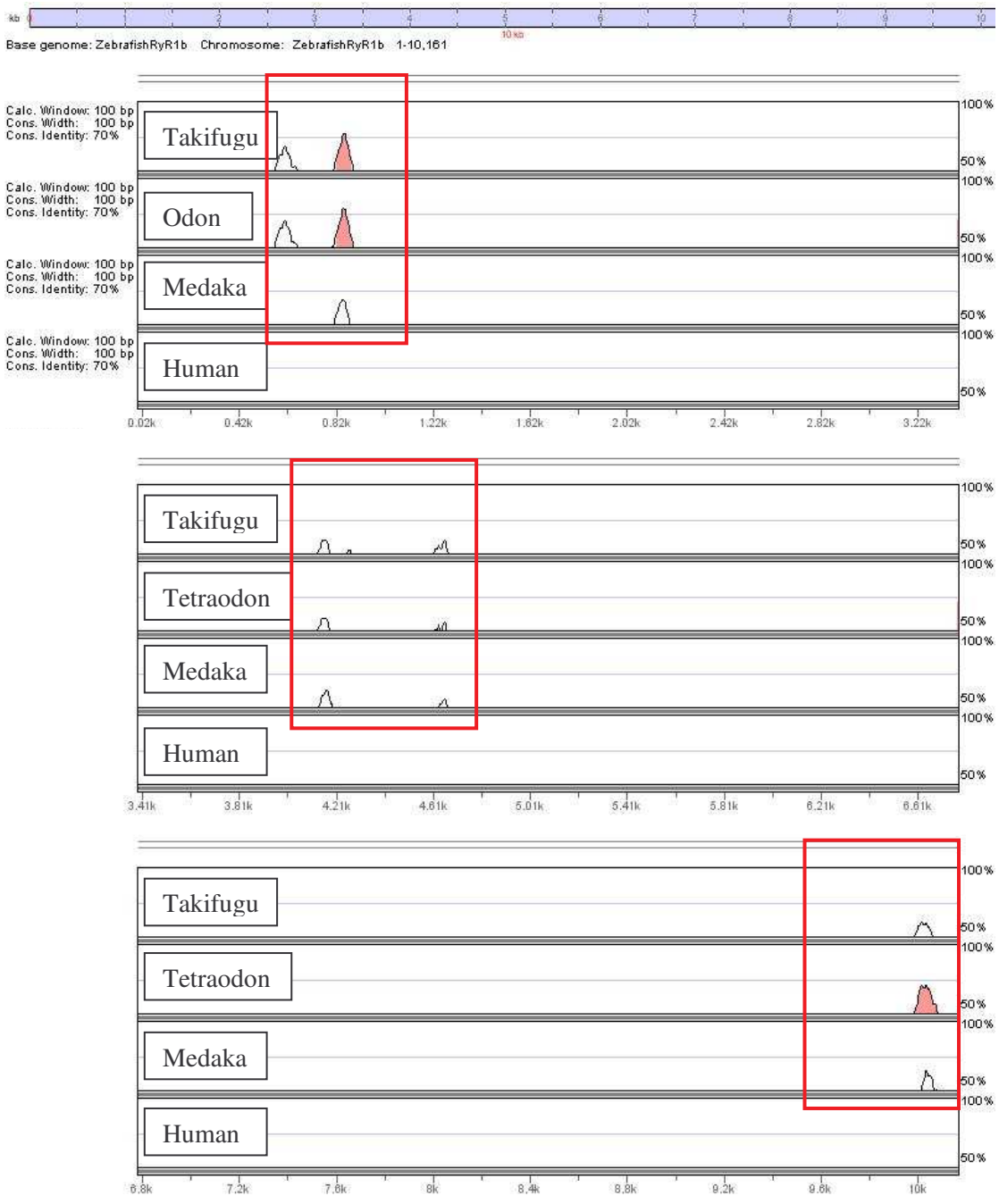
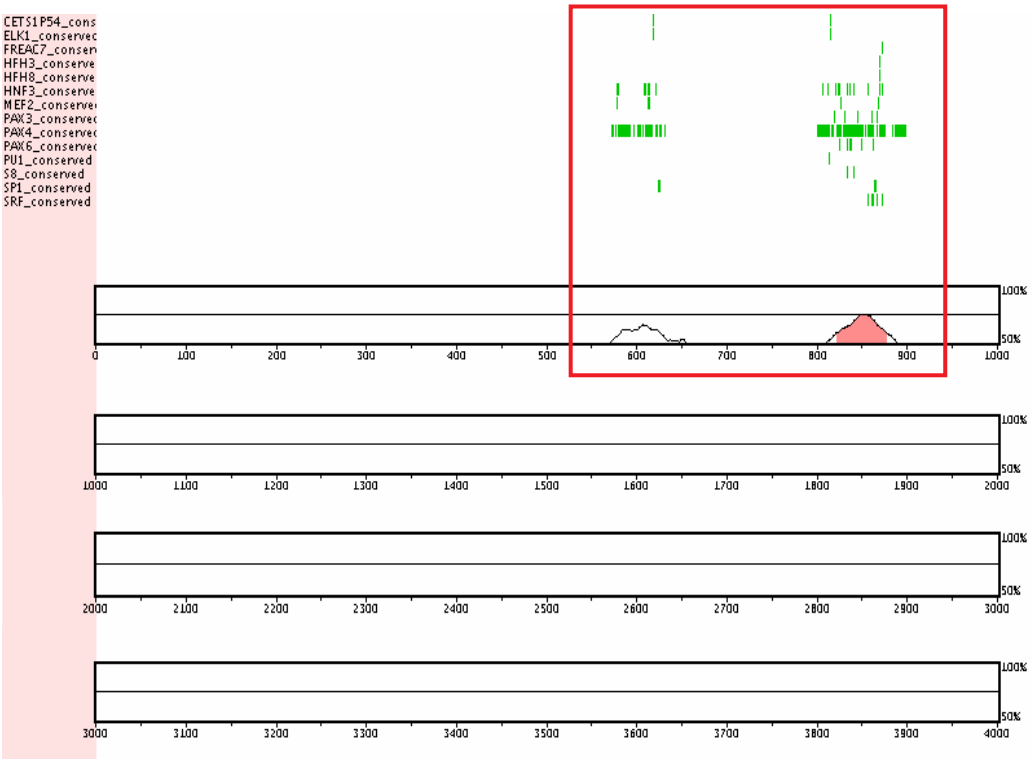


Figure 8. VISTA analysis: the zebrafish RyR1b promoter shows limited conservation with other teleost RyR1b promoters. Base sequence = zebrafish RyR1b upstream region. Base 0 corresponds to 10 kb upstream, base 10 161 is immediately adjacent to first exon. From top: alignment with upstream sequences of takifugu, tetraodon, medaka RyR1b-like genes and human RyR1. Alignment based on sliding 100 bp window; peaks indicate conservation above 50%, shading indicates conservation over at least 100 bp with minimum 70% base pair identity. Red boxes highlight conservation.

3.4 Conservation of TFBSs among Teleost RyR1a and RyR1b Promoters

The aligned sequences were next scanned for the matrices identified in Table 3 to identify conserved TFBSs. Not unexpectedly, none of these TFBSs were conserved between the aligned teleost upstream RyR1a sequences, scanned using a 24 bp sliding window at the recommended thresholds (one standard deviation below the average score of the true binding sites for each matrix). Conversely, for RyR1b, some conserved TFBSs were clustered in the three conserved sequence sites, including Sp1 and MEF2 in two of the three segments (Figure 9). Conserved sites can only be visualized between two organisms using VISTA; Figure 9 illustrates the zebrafish-takifugu alignment. Zebrafish-tetraodon and zebrafish-medaka yielded similar results, though zebrafish-tetraodon also shared some clustered TFBSs at -6100 and -1500.



CTD.

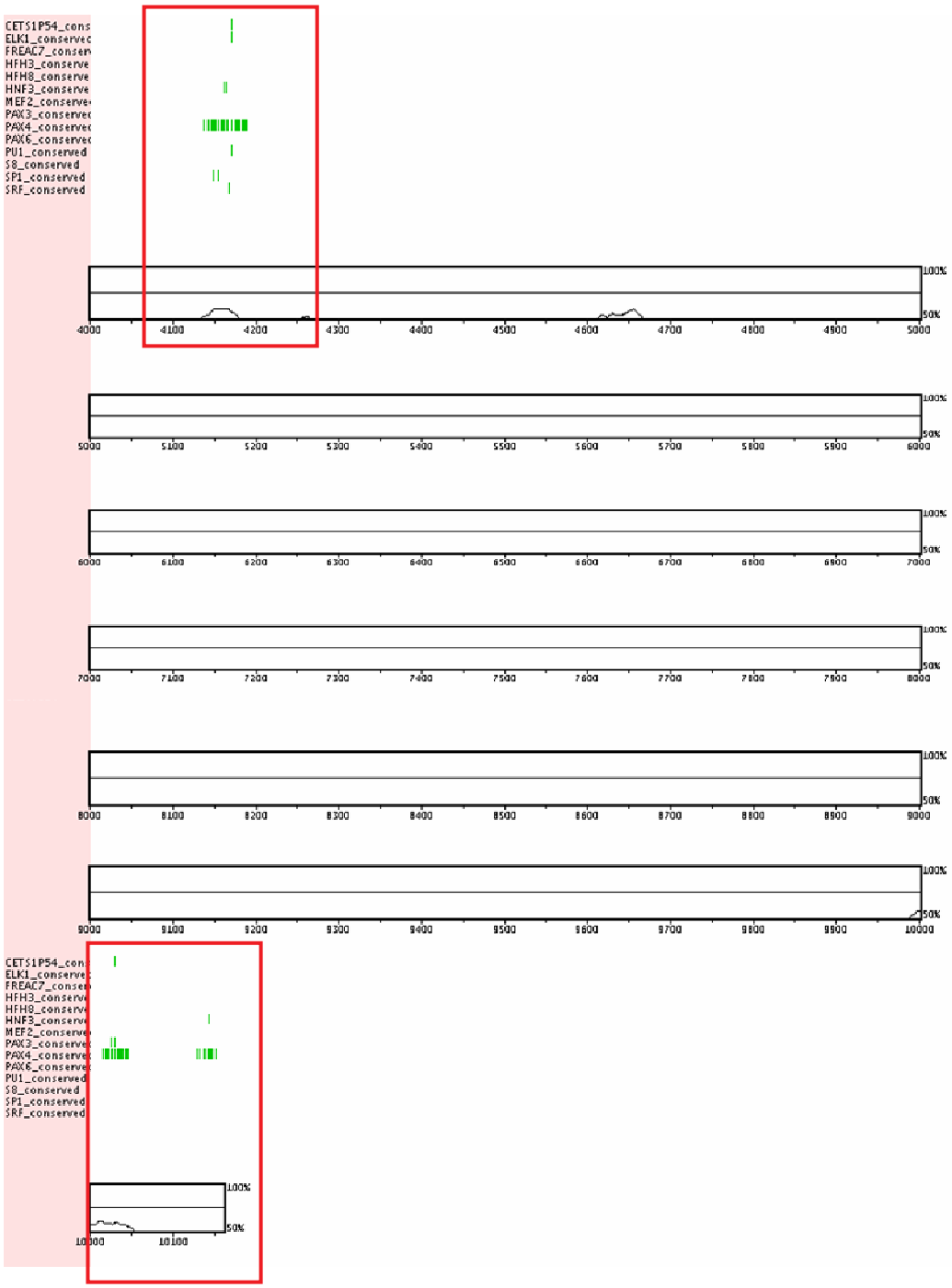


Figure 9. Some TFBSs are conserved between the zebrafish and takifugu 5' RyR1b sequences. Base sequence = zebrafish RyR1b upstream sequence. Base 0 corresponds to 10 kb upstream of the 5' UTR, base 10 181 is immediately adjacent to first exon. Illustrated is alignment with takifugu RyR1b-like upstream sequence. Conserved TFBSs indicated by green marks, identified by scanning with a 24 bp sliding window and recommended thresholds. Red boxes outline conserved clusters of TFBSs.

3.5 Sequence/TFBS Comparison between Zebrafish RyR1a and RyR1b Promoters

Comparison of zebrafish RyR1a and RyR1b promoters also indicated a lack of strong sequence conservation (Figure 10). However, several conserved muscle-specific TFBSs were identified between RyR1a and RyR1b, although they are all located very far upstream (Figure 11). Neither RyR1a nor RyR1b promoters share any conserved TFBSs with the human RyR1 promoter (plots not shown).

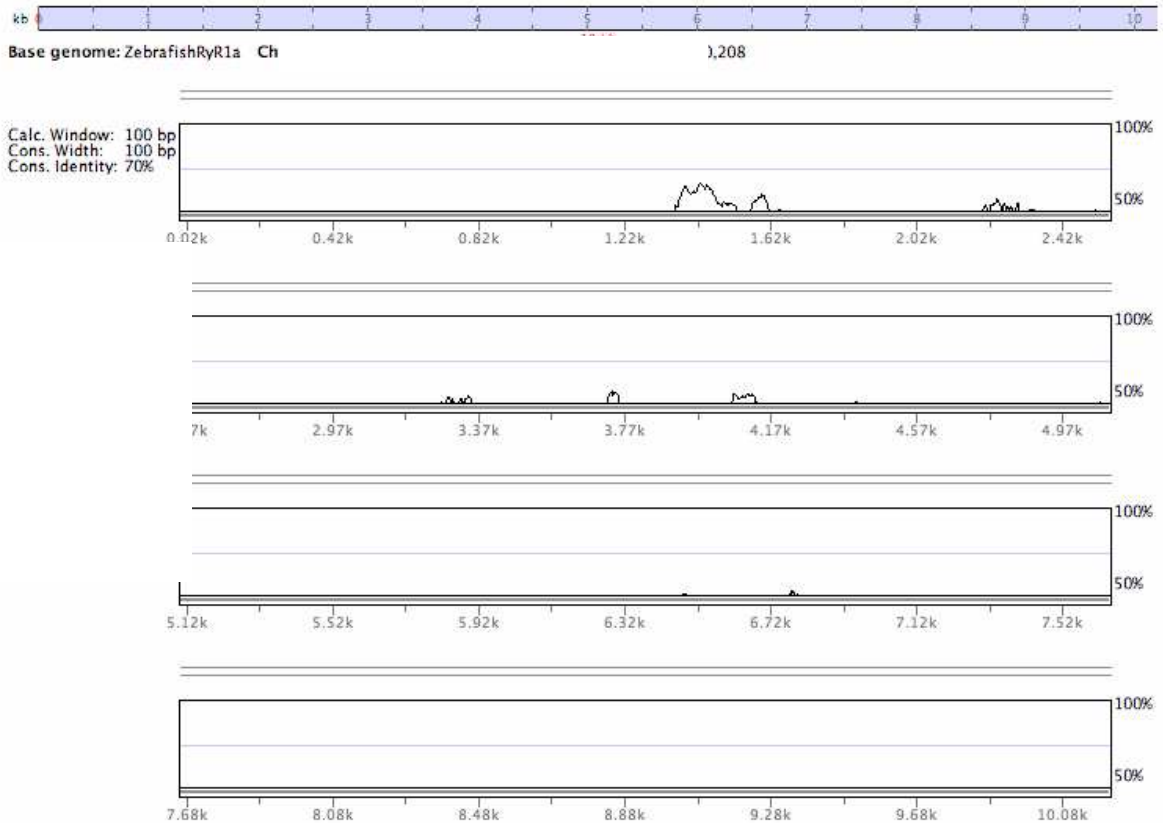


Figure 10. The zebrafish RyR1a and RyR1b promoters are not highly conserved. Base sequence = zebrafish RyR1a upstream sequence. Base 10 175 is immediately adjacent to first RyR1a exon. Illustrated is alignment with zebrafish RyR1b upstream sequence. Alignment based on sliding 100 bp window; peaks indicate regions conserved above 50% over 100 bp.

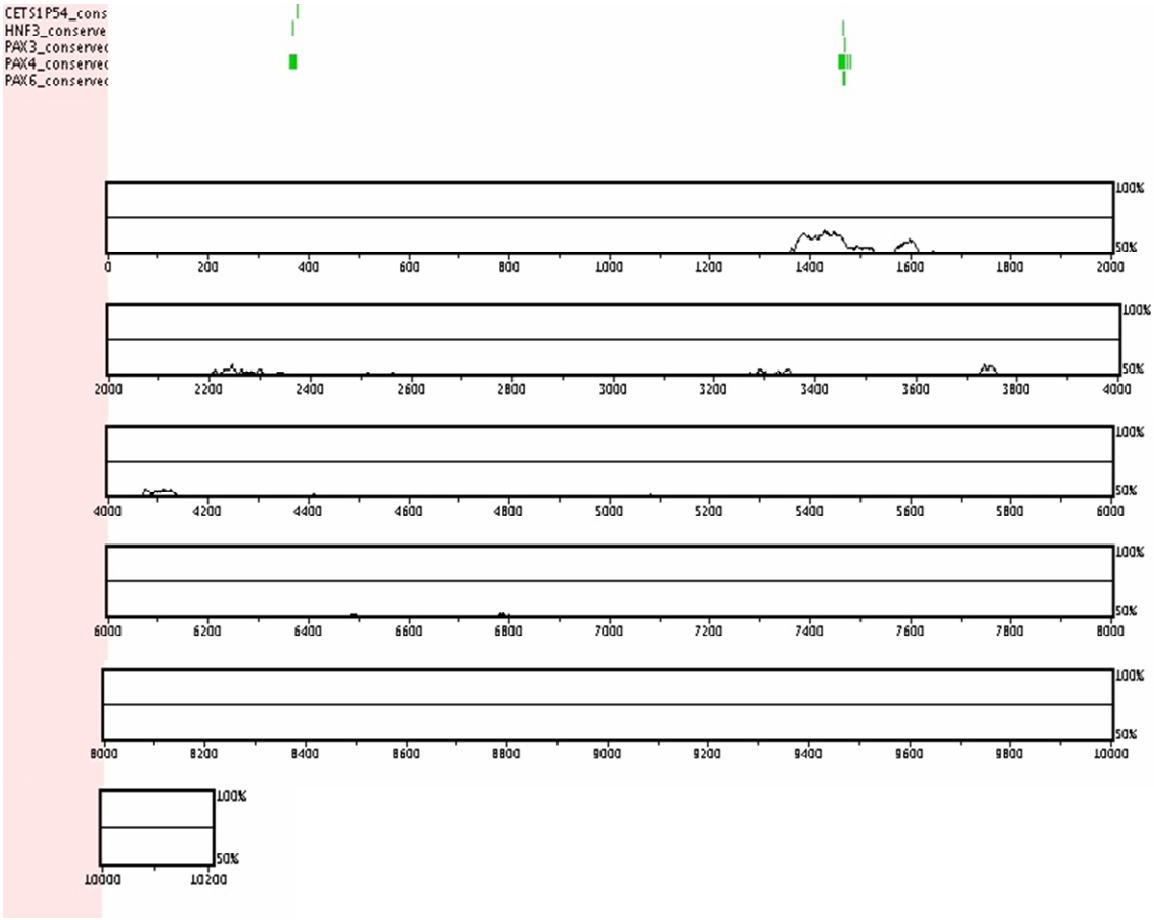


Figure 11. Some TFBSs are conserved between the zebrafish RyR1a and RyR1b sequences. Base sequence = zebrafish RyR1a upstream sequence. Base 0 corresponds to 10 kb upstream of the 5' UTR, base 10 175 is immediately adjacent to first exon. Illustrated is alignment with RyR1b upstream sequence. Conserved TFBS indicated by green bars, identified by scanning with a 24 bp sliding window and recommended thresholds.

3.6 RyR1a and RyR1b Proximal Promoter Sequence Analysis

In addition to *in silico* analysis of the RyR1a and RyR1b promoters, pGL4 expression vector promoter constructs were created to prepare for future functional analysis. In the process, sequencing (5-10-fold) of overlapping short cloned fragments of the RyR1a and RyR1b promoters verified the database 2000 bp proximal promoter sequences, with the exception of several polymorphisms within each. This is to be

expected because our zebrafish are likely of a different strain from those used to sequence and create the Zv8 assembly, and SNPs are frequent in the genome.

Two additional findings are of note. First, as pointed out in section 2.7, the promoters were again sequenced following subcloning into pGL4 to confirm identity of the polymorphic sites and the correct orientation. Sequencing of the RyR1a promoter at this point revealed a 201 bp insertion (from base -906 to -705 in Figure 12) that was not detected in the contig assembled from shorter sequenced fragments. Secondly, a microsatellite in the RyR1b promoter – $(GA)_4GC(GA)_{16}AG(GA)_6$ – was identified between bases -77 to -21 (Figure 13). Sequencing through the microsatellites isolated from multiple zebrafish genomic DNA revealed that this microsatellite is polymorphic (data not shown).

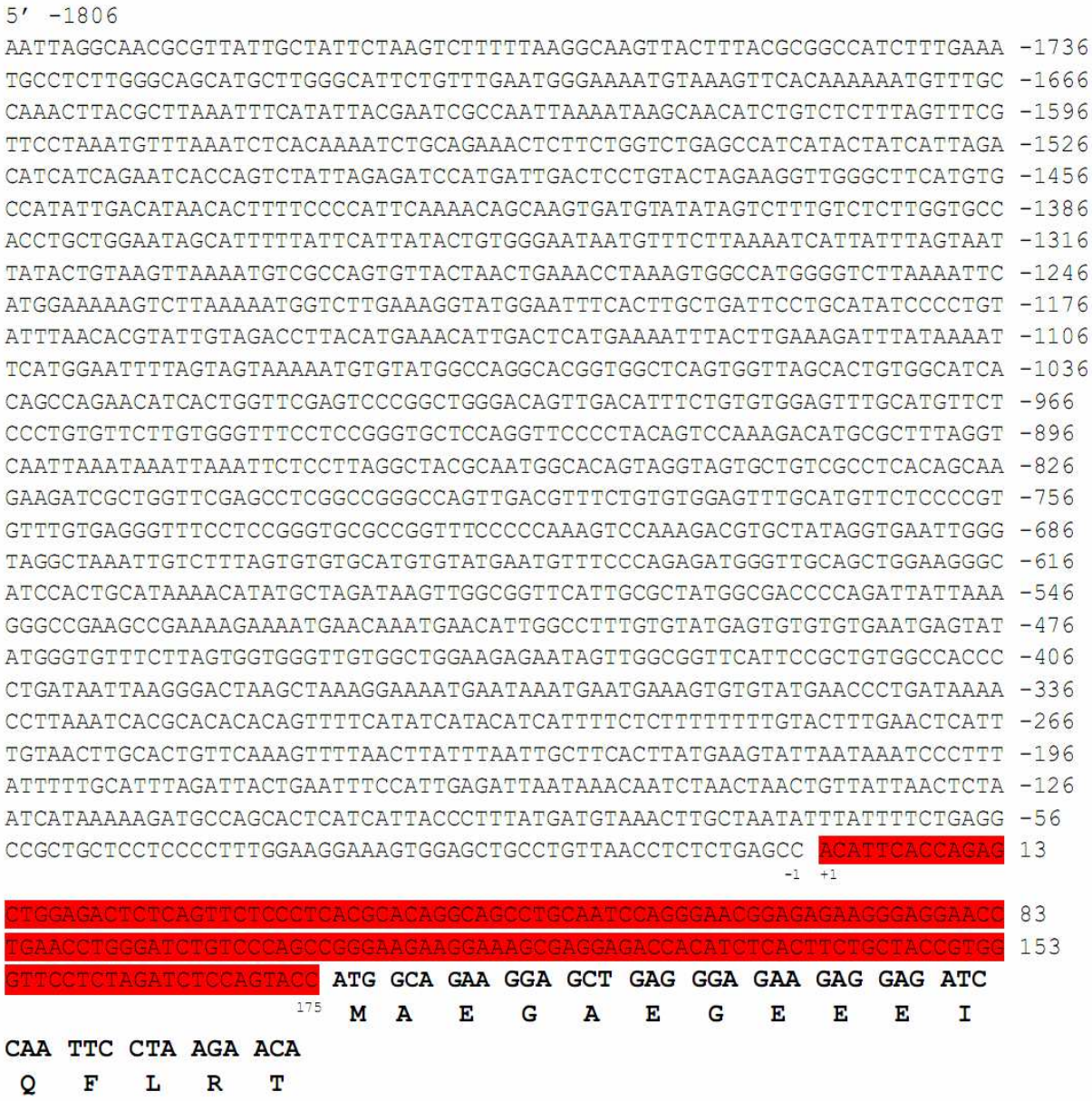


Figure 12. RyR1a proximal promoter as cloned into the pGL4 expression vector. Sequencing confirmed the correct sequence and orientation of the RyR1a proximal promoter in the pGL4 expression vector. Red shading identifies a 175 bp 5' UTR as identified by EST EB035351. The construct ends in codon 3 (E) but the sequence and amino acid translation of the entire first coding exon is included for reference. Several polymorphisms were identified as compared to the database sequence, as well as a 201 bp insert (see discussion).

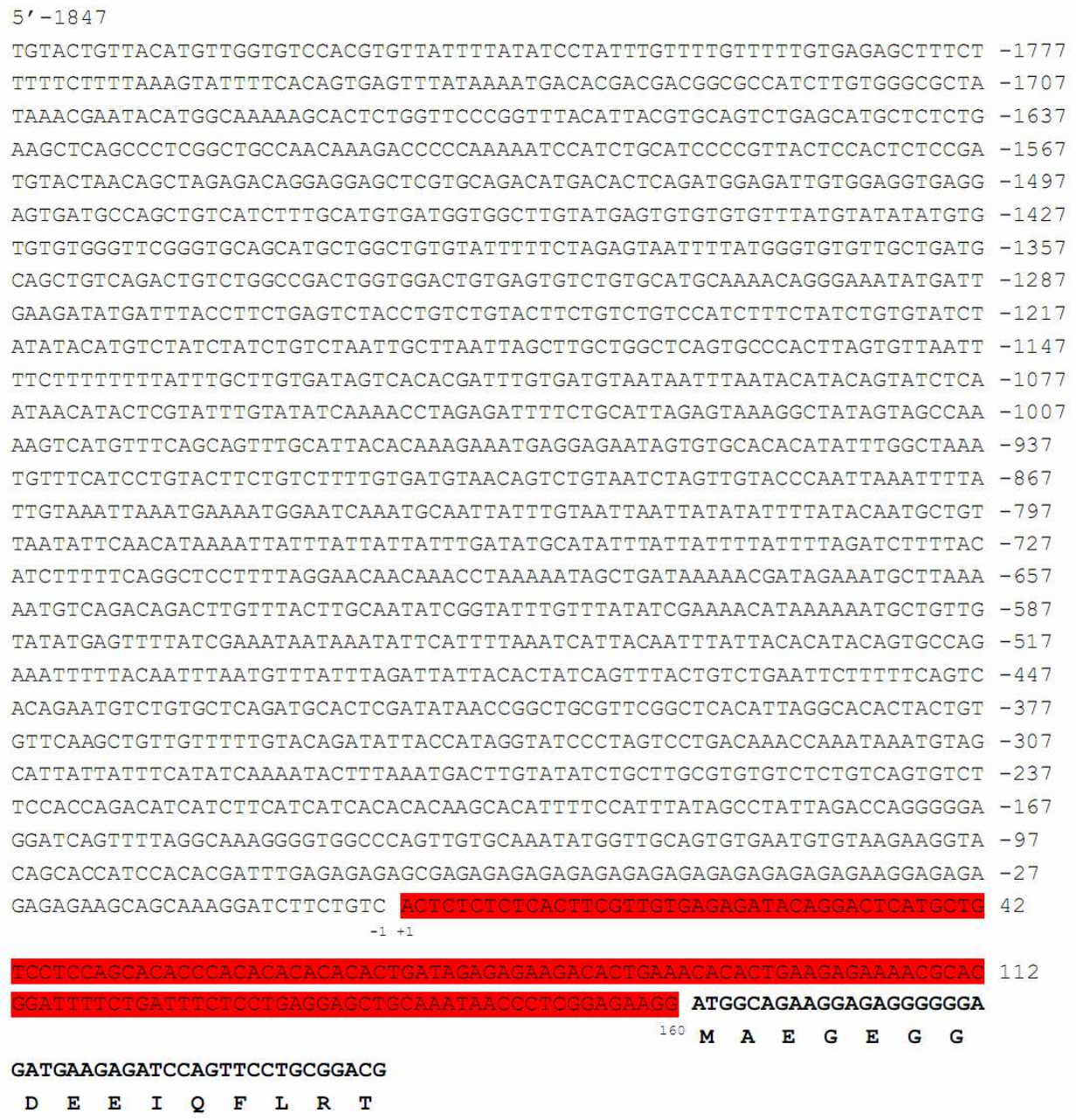


Figure 13. RyR1b proximal promoter as cloned into the pGL4 expression vector. Sequencing confirmed the correct sequence and orientation of the RyR1b proximal promoter in the pGL4 expression vector. Red shading identifies a 160 bp 5' UTR as identified by EST EB848566.1. The construct ends at +160 (just prior to the first amino acid codon) but the sequence and amino acid translation of the first coding exon is included for reference. Several polymorphisms were identified as compared to the database sequence.

4. Discussion

This project has set the foundation for a complete transcriptional analysis of the zebrafish RyR1a and RyR1b genes. An *in silico* analysis has successfully identified three putative regulatory elements in the RyR1b promoter. In addition, proximal promoter constructs have been created for future functional testing. Sequencing of these constructs revealed an insertion and microsatellite in the RyR1a and RyR1b promoters, respectively. An analysis of *cis*-regulatory divergence was performed by aligning the zebrafish promoters to each other and to comparative vertebrate species, including human; however, this analysis did not yield strong evidence to support or refute the mechanism.

4.1 Identification of Putative Regulatory Elements by *in silico* Analysis

This study revealed a considerable lack of sequence conservation and conserved TFBSs between RyR1a and RyR1b and their comparative vertebrate promoter sequences, particularly between RyR1a orthologues. This is intriguing since, as non-coding sequence conservation is being used as a sign of regulatory function, it implies that no regulatory elements are present in the zebrafish 5' RyR1a upstream sequence, despite many MatInspector hits. This is also surprising because expression of RyR1a is conserved in slow-twitch muscle among teleosts and transcriptional control is expected to be at least partially conserved.

There are two primary explanations for the deficiency of identified regulatory elements. First, important regulatory sequences are not solely located within the upstream sequence of a gene. While usually found in this region, transcription of genes can be controlled by enhancers and repressors located further upstream, downstream and even

within the genes themselves. For example, important *cis*-regulatory elements for the zebrafish *pax6a* and *pax6b* genes have been identified within the first intron of a downstream gene, ELP4 (Kleinjan *et al*. 2008). Further, muscle-specific expression of the porcine RyR1 gene was found to be regulated by a repressor in the first intron (Shmoelzl *et al*. 1996). Notably, the available zebrafish genome indicates that RyR1b has an extremely large ~55 kb first intron, which houses a CpG island (data not shown). This comparatively large first intron is also found in other teleost RyR1b genes. CpG islands are methylation-rich sites, often involved in regulating transcription (Gardiner-Garden and Frommer 1987), thus the first intron of the RyR1a and RyR1b genes should also be analyzed with comparative genomics tools.

Secondly, recent studies have emphasized the limitations of comparative sequence analysis in identifying regulatory regions (Fisher *et al*. 2006; McGaughey *et al*. 2005). To test this oft-used criterion, Fisher *et al*. (2006) compared the genomic sequences upstream of the *ret* receptor tyrosine kinase gene in zebrafish and *RET* in human. These regions are not conserved between zebrafish and human; however, some sequences are conserved among teleosts and some among mammals. The ability of these sequences to drive transcription was functionally analyzed; the authors found that 11 out of 13 mammalian conserved sequences were able to drive zebrafish-specific transcription in a Tol2-transposon based assay in zebrafish embryos. This indicates that regulatory function is conserved between human and zebrafish *ret/RET* loci, despite the lack of sequence similarity.

A more recent combined bioinformatics-experimental study by McGaughey and colleagues (2008) found that established comparative genomics programs only recognized 29-61% of functional regulatory regions experimentally located within the *phox2b* promoter. As in the previous study, and consistent with prior anecdotal results, the authors concluded that locating regulatory elements by sequence constraint alone will overlook a significant amount of regulatory information. Thus, the 10 kb sequence here analyzed may contain regulatory elements missed by the limitations of current bioinformatics tools.

Despite these limitations, as noted earlier, some regulatory information was detected in the RyR1b upstream sequence. The putative zebrafish RyR1b regulatory regions are particularly notable for two reasons. First, one of these regions is located within the 5' UTR in the created pGL4 construct, leading to the hypothesis that this construct will successfully drive transcription. Secondly, Sp1 and MEF2 are among the conserved TFBSs in two of the three conserved regions. This is interesting due to the importance of these same TFBSs in the porcine RyR1 promoter (Schmoelzl *et al*. 1996).

4.2 *cis*-Regulatory Divergence

A secondary goal of this project was to begin an analysis of *cis*-regulatory divergence between the RyR1a and RyR1b promoters. Here, it has been identified that the RyR1a and RyR1b upstream sequences are not highly conserved, which would be expected under a model of *cis*-regulatory divergence; however, this does not strongly support *cis*-regulatory divergence (Kleinjan *et al*. 2008). Further, as no putative regulatory elements were associated with RyR1a and no segment of sequence for either

RyR1a nor RyR1b were conserved with the ancestral human sequence, it is not possible to identify whether ancestral regulatory elements have been divided between the two paralogues. Using sequence conservation as a marker of regulatory function has in turn limited this analysis.

4.3 Sequence Analysis of the RyR1a and RyR1b Proximal Promoters

As mentioned, sequencing of the pGL4 constructs revealed that the majority of both the RyR1a and RyR1b sequences are nearly identical to the latest respective database sequences except for several haplotypic sites each. However, a 201 bp insertion was identified within the RyR1a-pGL4 promoter construct that was not detected in shorter amplified PCR product sequence reads. Since multiple genomic extractions were used, it is possible that the RyR1a promoter is polymorphic for the insertion. Alternatively, the promoter and some of the RyR1a gene could be duplicated within the genome, with one copy gaining this insertion. Supporting the former hypothesis is that in Zv7, the previous zebrafish genome assembly, the upstream sequence of RyR1a did contain the insertion and the sequences in Zv7 and Zv8 are otherwise nearly identical. Unless the duplication was very recent, there should be significant changes accumulated between the two copies due to lack of selective pressure in non-regulatory, non-coding sequence (Bird *et al*. 2006). Further, only one PCR product was obtained when the 2000 bp promoter was amplified, suggesting there is only one copy in the genome. Regardless, the effect of the insertion on the ability of the promoter to drive transcription warrants further investigation.

Another interesting finding in the sequence of the proximal promoters is the RyR1b microsatellite. Other promoter analyses have identified functionally important microsatellites, where the rate of transcription depends on the length of the microsatellite (Bormann *et al*. 2003; Wang *et al*. 2005). For example, the human HMGA2 promoter includes a (TCTCT$(TC)_n$) repeat; Bormann and colleagues (2003) found that luciferase vectors containing a single TC repeat had only 55% promoter activity compared to a 22-repeat clone, and that the 22-repeat clones had only 84% promoter activity compared to a 36-repeat clone. Thus, the polymorphism of the RyR1b microsatellite is another area for further exploration.

4.4 Future Work

Further *in silico* analysis will be required to identify putative regulatory elements, particularly for the RyR1a promoter. First, the first introns will be analyzed in a manner similar to the protocol used in this study. In addition, other promoter prediction programs will be tested for their ability to detect regulatory elements in the RyR1a and RyR1b 5' sequences. New computational programs are constantly being developed and upgraded due to the recent push to identify regulatory information in the non-coding genome; these programs are increasing in accuracy as more experimental data is added to databases. Improved position weight matrices to more sensitively identify TFBSs, known clusters of conserved TFBSs, chromatin structure and ideal TF concentration are all factors now being incorporated alongside non-coding sequence conservation to detect regulatory elements (Narlikar and Ovcharenko 2009).

The promoter-pGL4 constructs created as described will next be tested for their ability to drive transcription. Based on the limited information presented here, it is predicted that the RyR1a construct promoter activity will be null, while the RyR1b construct will drive transcription at some level due to the predicted regulatory element and the microsatellite. Transient expression arrays will be carried out by injection of the promoter-pGL4 constructs into zebrafish tissue alongside pRenilla plasmids, which constitutively express a distinguishable bioluminescent protein (*Renilla* luciferase) to serve as a baseline for each injection procedure (Swanson *et al*. 2007). The respective levels of expression of firefly luciferase and *Renilla* luciferase will be compared to determine the ability of the promoters to drive expression.

Since it is known that RyR1a and RyR1b are expressed in skeletal muscle, the ideal method is microinjection into zebrafish skeletal muscle. Injection into adult fish muscle has been successfully performed with other gene promoter constructs (Schulte and Powers 1998). This system will not only determine whether the constructs are able to drive muscle transcription, but could also potentially determine whether the 2000 bp promoters are sufficient to drive fast and slow-twitch specific expression.

There is potential for further functional analysis regardless of whether or not these constructs are able to drive general or muscle-specific transcription. If successful, the 2000 bp sequences will be dissected to determine which segments are required for transcription, by deleting portions and testing the ability of the remaining sequence to drive transcription. For RyR1b, if transcription is successful, the logical first deletions will be the conserved region and the microsatellite, with the prediction that without one

or both of these elements, promoter activity will decrease. If, in contrast, the proximal promoters are not able to drive transcription, alternative sequences can be tested; for example, a proximal promoter lacking the 201 bp insertion, the other two putative regulatory elements in the RyR1b 10 kb upstream sequence and any predicted regulatory elements identified in the first intron or by further *in silico* analyses.

More bioinformatics and functional work must be done, on both ancestral RyR1 and zebrafish RyR1a and RyR1b promoters, before clear evidence for *cis*-regulatory divergence will be found. In carrying out further analysis, two factors will be kept in mind. First, though human RyR1 is acceptable for a preliminary analysis, it is not the best ancestral sequence for comparison: an RyR1 promoter closer in evolutionary sequence to the RyR1a and RyR1b duplication would be ideal, or several mammalian sequences should be examined. It is also worth noting that, by comparative synteny analysis, it has been identified that the RyR1b gene shares neighbouring genes with the human RyR1 gene while the RyR1a gene does not, suggesting RyR1b has endured more sequence constraint (Franck *et al*. in preparation). I predict that further analysis will reveal that the RyR1b gene shares more conserved regulatory elements with the ancestral RyR1 gene than does RyR1a.

5. Conclusions

1. The zebrafish RyR1a 10 kb upstream sequence is not conserved with other RyR1a upstream sequences, suggesting a lack of regulatory elements.

2. The zebrafish RyR1b 10 kb upstream sequence is not highly conserved with other RyR1b upstream sequences, though there are several sites which show 70% conservation between teleosts, and one is over 100 bp long. There are identified conserved TFBSs within these regions, suggesting three candidate regulatory regions.

3. The zebrafish RyR1a and RyR1b 10 kb upstream sequences are not highly conserved and neither share any sequence conservation with human RyR1 upstream sequence. Functional and improved bioinformatics analyses must be carried out before definitive evidence of *cis*-regulatory divergence will be identified.

4. Using sequence conservation as an indicator of regulatory function is useful but likely misses a significant amount of regulatory information.

5. The RyR1a promoter is polymorphic for a 201 bp insert.

6. There is a polymorphic microsatellite in the RyR1b promoter just before the predicted transcription start site.

7. Further work, including transient expression arrays of the created proximal promoter constructs, creation of deletion constructs and analysis of the first introns, is ready to be performed.

6. References

Bird, C.P., Stranger, B.E. and E.T. Dermitzakis. 2006. Functional variation and evolution of non-coding DNA. *Curr Opin Genet Dev* 16: 559-564.

Bormann, L., Seebeck, B., Rogalla, P. and J. Bullerdieck. 2003. Human *HMGA2* promoter is coregulated by a polymorphic dinucleotide (TC)-repeat. *Oncogene* 22: 756-760.

Carrtharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M. and T. Werner. 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* 21: 2933-2942.

Copello, J.A., Zima, A.V., Diaz-Sylvester, P.L., Fill, M. and L.A. Blatter. 2007. Ca^{2+} entry-independent effects of L-type Ca^{2+} channel modulators on Ca^{2+} sparks in ventricular myocytes. *AJP Cell Physiol* 292: 2129-2140.

Darbandi, S. and J.P.C. Franck. 2009. A comparative study of ryanodine receptor (RyR) gene expression levels in a basal ray-finned fish, bichir (*Polypterus ornatipinnis*) and the derived euteleost zebrafish (*Danio rerio*). *Comp Biochem and Physiol B* 154: 443-448.

Dunwoodie, S.L., Joya, J.E., Arkell, R.M. and E.C. Hardeman.1994. Multiple regions of the human cardiac actin gene are necessary for maturation-based expression in striated muscle. *J Biol Chem* 269 (16): 12212-12219.

Fill, M. and J. Copello. 2002. Ryanodine receptor calcium release channels. *Physiol Rev* 82 (4): 893-9

Fisher, S., Grice, E.A., Vinton, R.M. and S.L. Bessling. 2006. Conservation of *RET* regulatory function from human to zebrafish without sequence similarity. *Science* 312: 276-279.

Franck, J., Morisette, J., Keen, J., Londraville, R., Beamsley, M. and B. Block. 1998. Cloning and characterization of fiber-type specific ryanodine receptor isoforms in skeletal muscles of fish. *AJP Cell Physiol* 275 (2): 401-415.

Franck, J.P.C., Reimer, K. and K. Hill. 2010. Molecular evolution of the ryanodine receptor gene family: a phylogenetic and comparative genomic analysis. Manuscript in preparation.

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M. and I. Dubchak. 2005. VISTA: computational tools for comparative genomics. *Nucl Acids Res* 32: 273-279.

Gardiner-Garden, M. and M. Frommer. 1987. CpG islands in vertebrate genomes. *J Mol Biol* 196: 261-282.

Hirata, H., Watanabe, T., Hatakeyama, J., Sprague, S., Saint-Amant, L., Nagashima, A., Cui, W., Zhou, W and J. Kuwada. 2007. Zebrafish *relatively relaxed* mutants have a ryanodine receptor defect, show slow swimming and provide a model of multi-minicore disease. *Development* 134: 2771-2781.

Ingham, P. 2009. The power of the zebrafish for disease analysis. *Hum Mol Genet* 18: R107-R112.

Jegga, A.G., Sherwood, S.P., Carman, J.W., Pinski, A.T., Phillips, J.L., Pestian, J.P. and B.J. Aronow. 2002. Detection and visualization of compositionally similar *cis*-regulatory element clusters in orthologous and co-ordinately controlled genes. *Genome Res* 12: 1408-1417.

Kleinjan, D., Bancewicz, R., Gautier, P., Dahm, R., Schonthaler, H., Damante, G., Seawright, A., Hever, A., Yeyati, P., van Heyningen, V. and P Coutinho. 2008. Subfunctionalization of duplicated zebrafish *pax6* genes by *cis*-regulatory divergence. *PLoS Genetics* 4 (2): 1-14.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and D.G. Higgins. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947-2948.

MacCarthy, T. and A. Bergman. 2007. The limits of subfunctionalization. *BMC Evol Biol* 7: 213-227.

McCarthy, T.V., Quane, K.A. and P.J. Lynch. 2000. Ryanodine receptor mutations in malignant hyperthermia and central core disease. *Hum Mutat* 15 (5): 410-417.

McGaughey, D.M., Vinton, R.M., Huynh, J., Al-Saif, A., Beer, M.A. and A.S. McCallion. 2008. Metrics of sequence constraint overlook regulatory sequences in an exhaustive analysis at *phox2b*. *Genome Res* 18:252-260.

Meyer, A. and Y. van de Peer. 2005. From 2R to 3R: evidence for a fish-specific genome duplication. *BioEssays* 27: 937-945.

Narlikar, L. and I. Ovcharenko. 2009. Identifying regulatory elements in eukaryotic genomes. *Brief Funct Genomic Proteomic* 8: 215-230.

Ntshida, K., Otsu, K., Hori, M., Kuzuya, T. and M. Tada. 1996. Cloning and characterization of the 5'-upstream regulatory region of the Ca^{2+}-release channel gene of cardiac sarcoplasmic reticulum. *Eur J Biochem* 240: 408-415.

Otsu, K., Willard, H.F., Khanna, V.K., Zorzato, F., Green, N.M. and D.H. MacLennan. 1990. Molecular cloning of cDNA encoding the Ca^{2+} release channel (ryanodine receptor) of rabbit cardiac muscle sarcoplasmic reticulum. *J. Biol. Chem* 265: 13472-13483.

Phillips, M., Fujii, J., Khanna, V., DeLeon, S., Yokobata, K., de Jong, P. and D. MacLennan. 1996. The structural organization of the human skeletal muscle ryanodine receptor (RyR1) gene. *Genomics* 34: 24-41.

Rozen, S. and H.J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S., and S. Misener (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology.* Humana Press: Totowa, NJ, 365-386.

Schmoelzl, S., Leeb, T., Brinkmeier, H., Brem, G. and B. Brenig. 1996. Regulation of tissue-specific expression of the skeletal muscle ryanodine receptor gene. *J Biol Chem* 271 (9): 4763-4769.

Schulte, P.M. and D.A. Powers. 1998. Efficient gene transfer into *Xiphophorus* muscle and melanoma by injection of supercoiled plasmid DNA. *Mol Mar Biol and Biotech* 7 (4): 241-247.

Sorrentino, V., Barone, V. and D. Rossi. 2000. Intracellular Ca^{2+} calcium release channels in evolution. *Curr Opin Genet Dev* 10: 662-667.

Swanson, B., Fan, F. and K. Wood. 2007. Enhanced response dynamics for transcription analysis using new pGL4 luciferase reporter vectors. *Cell Notes* 17: 3-5.

Takeshima, H., Nishimura, S., Matsumoto, T., Ishida, H., Kangawa, K., Minamino, N., Matsuo, H., Ueda, M., Hanoaka, M., Hirose, T. and S. Numa. 1989. Primary structure and expression from complementary DNA of skeletal muscle ryanodine receptor. *Nature* 339: 339-345.

Vincze, T., Posfai, J. and R.J. Roberts. 2003. NEBcutter: a program to cleave DNA with restriction enzymes. *Nucl Acids Res* 31: 3688-3691.

Wang, B., Ren, J., Ooi, L., Chong, S. and C. Lee. 2005. Dinucleotide repeats negatively modulate the promoter activity of Cyr61 and is unstable in hepatocellular carcinoma patients. *Oncogene* 24: 3999-4008.

Yang, C., Bolotin, E., Jiang, T., Sladek, S.M. and E. Martinez. 2007. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* 389: 52-65.

Zorzato, F., Fujii, J., Otsu, K., Phillips, M., Green, N.M., Lai, F.A., Meissner, G. and D.H. MacLennan. 1990. Molecular cloning of cDNA encoding human and rabbit forms of the Ca^{2+} release channel (ryanodine receptor) of skeletal muscle sarcoplasmic reticulum. *J Biol Chem* 265: 2244-2256.

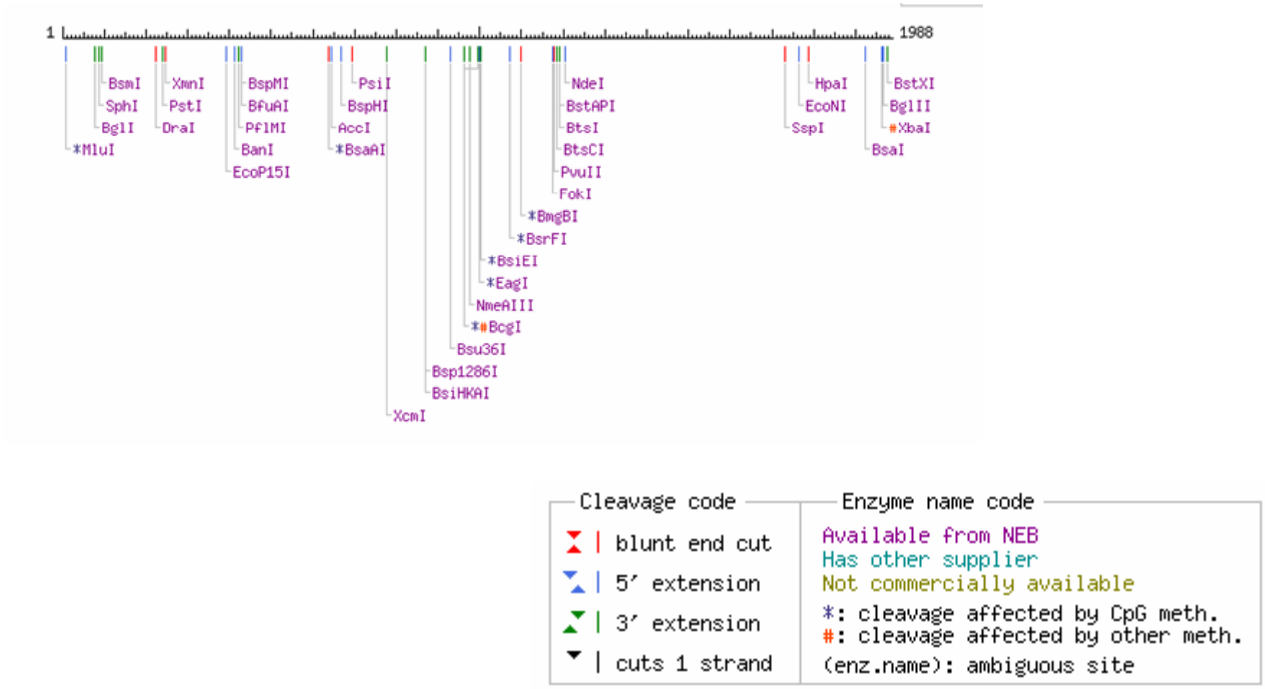
7. Appendices

7.1 Transcription Factor Information (adapted from MatInspector). Listed are the transcription factor families with skeletal muscle expression and individual matrices within each family. Optimized matrix similarity = degree of similarity required for a positive hit, calculated using previous data for the length of each binding site.

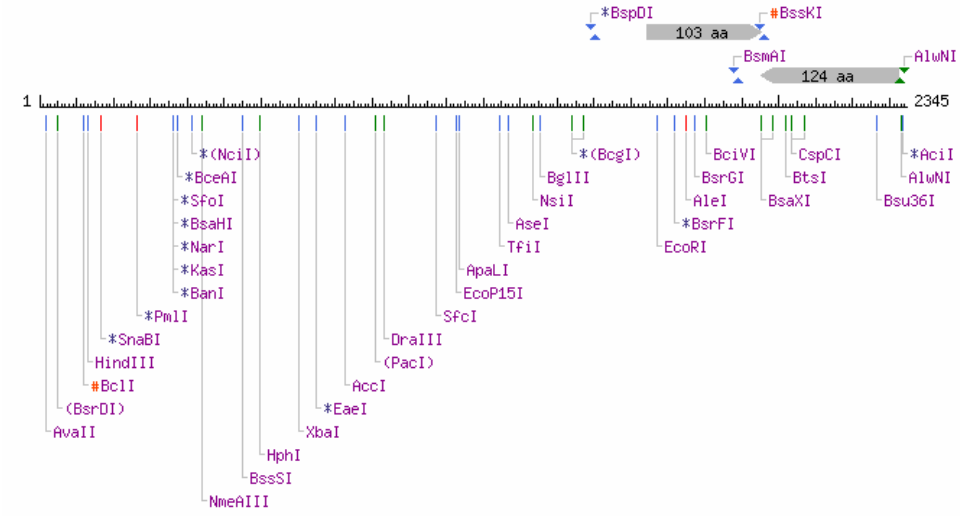
Family	Family Information	Matrix Name	Information	Optimized matrix sim.
FKHD	Fork head domain factors	FREAC2	Fork head related activator-2 (FOXF2)	0.84
		FREAC3	Fork head related activator-3 (FOXC1)	0.84
		FREAC4	Fork head related activator-4 (FOXD1)	0.78
		FREAC7	Fork head related activator-7 (FOXL1)	0.96
		HFH1	HNF-3/Fkh Homolog 1 (FOXQ1)	0.84
		HFH3	HNF-3/Fkh Homolog 3 (FOXI1, Freac-6)	0.97
		HFH8	HNF-3/Fkh Homolog-8 (FOXF1)	0.92
		HNF3	Hepatocyte nuclear factor 3 (alpha, beta) (FOXA1, FOXA2)	0.98
		XFD1	Xenopus fork head domain factor 1 (FoxA4a)	0.90
		XFD2	Xenopus fork head domain factor 2 (FoxI1a)	0.89
		XFD3	Xenopus fork head domain factor 3 (FoxA2a)	0.82
CART	Cart-1 (cartilage homeoprotein 1)	ALX4	Aristaless-like homeobox 4	0.82
		CART1	Cart-1 (cartilage homeoprotein 1)	0.86
		S8	Binding site for S8 type homeodomains	0.97
PAXH	PAX homeodomain binding sites	PAX4	Paired box 4, homeodomain binding site	0.84
		PAX6	Paired box 6, homeodomain binding site	0.86
TEAF	TEA/ATTS DNA binding domain factors	TEF1	TEF-1 related muscle factor	0.84
ETSF	Human and murine ETS1 factors	CETS1P54	c-Ets-1(p54)	0.92
		ELK1	Elk-1	0.81
		ETS1	c-Ets-1 binding site	0.92
		ETS2	c-Ets-2 binding site	0.84
		GABP	GABP: GA binding protein	0.86
		NRF2	Nuclear respiratory factor 2	0.86
		PEA3	Polyomavirus enhancer A binding protein 3, ETV4 (Ets variant gene 4)	0.94
		PU1	Pu.1 (Pu120) Ets-like transcription factor	0.89
SRFF	Serum response element binding factor	SRF	Serum response factor	0.84
MEF2	MEF2, myocyte-specific enhancer binding factor	MEF2	Myocyte-specific enhancer factor 2	0.89
		RSRFC4	Related to serum response factor, C4	0.85
MEF3	MEF3 binding sites	MEF3	MEF3 binding site, present in skeletal muscle-specific transcriptional enhancers	0.89
MYOD	Myoblast determining factors	E47	MyoD/E47 and MyoD/E12 dimers	0.97
		MYOD	Myogenic regulatory factor MyoD (myf3)	0.88
		MYO-GENIN	Myogenic bHLH protein myogenin (myf4)	0.91
PAX3	PAX-3 binding sites	PAX3	Pax-3 paired domain protein, expressed in embryogenesis, mutations correlate to Waardenburg Syndrome	0.76
SP1	GC-Box factors SP1/GC	GC	GC box elements	0.88
		SP1	Stimulating protein 1, ubiquitous zinc finger TF	0.88

7.2 Proximal Promoter (2000 bp) Restriction Enzyme Maps. Maps generated using NEBCutter for the a) RyR1a and b) RyR1b promoter regions.

a)

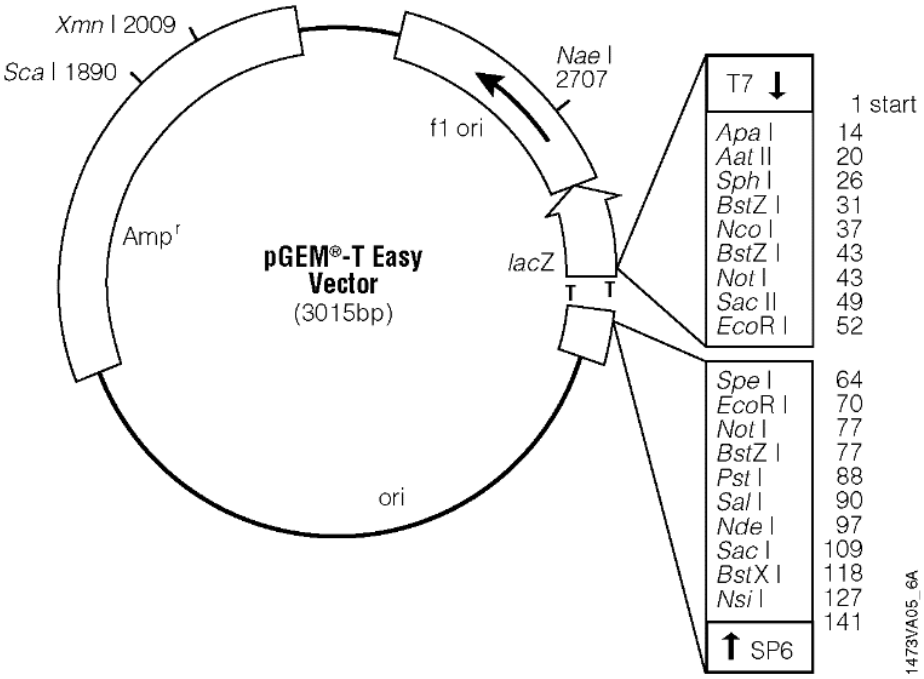


b)

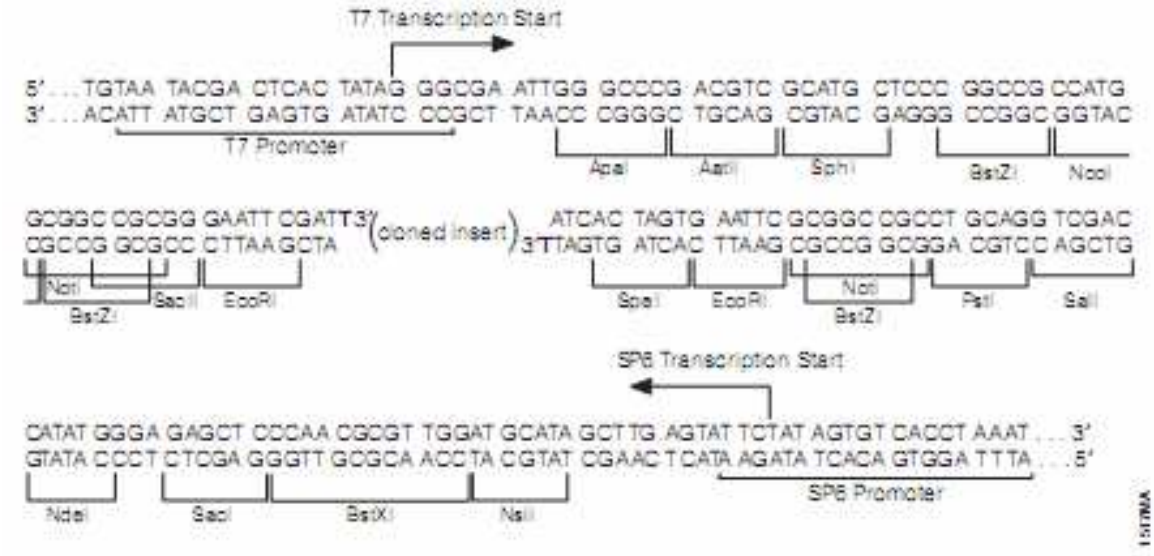


7.3 pGEM-T Easy Vector Maps. a) Schematic of vector, showing lacZ marker gene, ampicillin resistance gene and site of insertion of PCR products (T overhangs). b) Sequence flanking insertion site.

a)

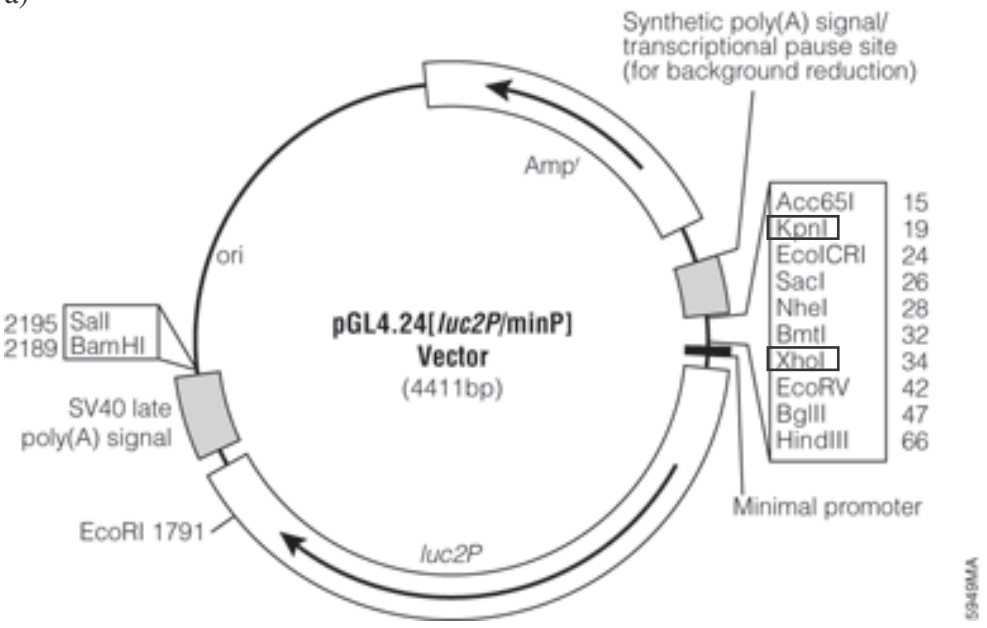


b)

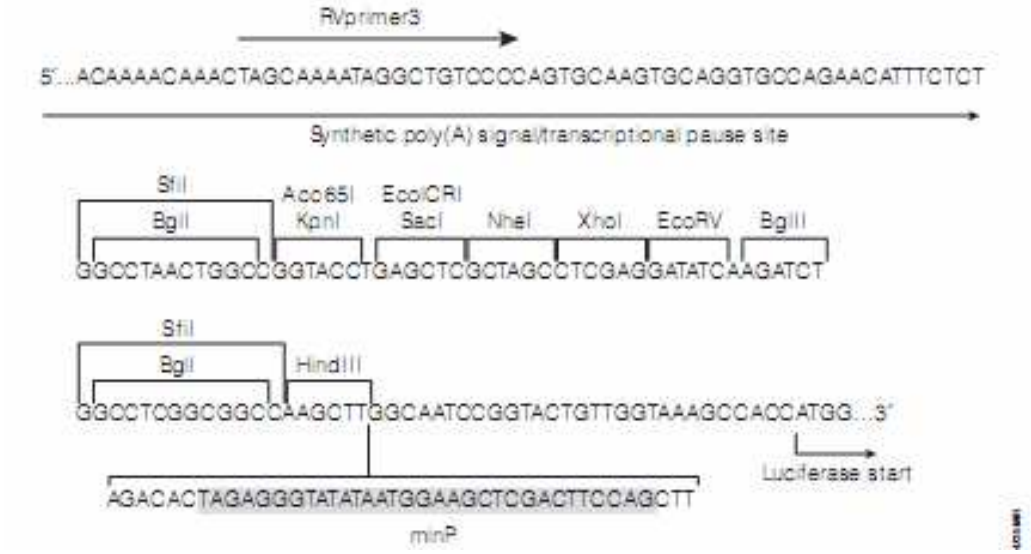


7.4 pGL4 Vector Maps. a) Schematic showing ampicillin resistance gene, luciferase gene (luc2P) and MCS upstream of luc2P with the KpnI and XhoI restriction enzyme sites selected for insertion of promoter. b) sequence of MCS; RV3 upstream primer is indicated.

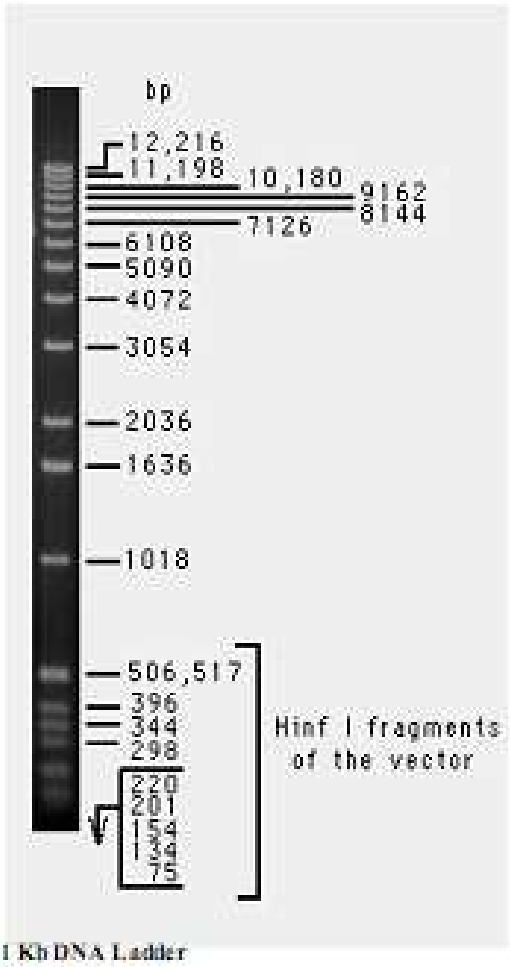
a)



b)



7.5 Ladder (1 kb)



1 Kb DNA Ladder
0.5 µg/lane
0.9% agarose gel
stained with ethidium bromide