Getting Promoted While Moving Down the Ladder? Characterization of the 5' Upstream Promoter Region of Ryanodine Receptor 3 (RyR3) in Zebrafish, *Danio rerio*

Ian Kasloff

Supervisor: Dr. Jens Franck

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

Department of Biology

The University of Winnipeg

2010

Abstract

Calcium plays a critical role in development, nerve transmission, and muscle contraction. The ryanodine receptor (RyR) is responsible for controlled release of calcium from the endoplasmic reticulum (ER) in neurons and the sarcoplasmic reticulum (SR) in muscle cells. Of the known RyR genes in fish, the RyR3 calcium release channel may play a more significant role in development as it is expressed in wide variety of tissues, including the brain and is upregulated during embryonic development. Currently, little is known about regulation of RyR3. A 2 kb sequence identified to be upstream of RyR3 was amplified and subcloned into an expression vector from genomic DNA isolated from zebrafish, *Danio rerio* and a contig of this sequence was built. An additional 8 kb was identified from the database for a total of 10 kb upstream of coding sequence. This sequence was analyzed for transcription factor binding sites (TFBS) using three programs; MatInspector, Match, and rVISTA. Comparisons to RyR3 promoter regions in three other fish species and humans show little sequence conservation. An abundance of hits was found for developmental factors such as HNF-1 and OCT-1, while relatively few sites for muscle-specific TFs were discovered. These results suggest RyR3 may have a role in the development of both contractile and non-contractile tissues, and that regulation of RyR3 may involve the complex interaction of several factors. Functional studies are required to confirm this inference.

Acknowledgments

When I first met my supervisor Dr. Jens Franck, I was a third year undergraduate student looking for inspiration. It was his work on Ryanodine receptor characterization that provided me with such motivation to pursue research in the field of molecular cell biology. It was his dedication to my project, coupled with his endless patience for my questions, which helped me become confident in molecular biology research. While it is never completely possible to express my gratitude within the confinements of this page, I hope he is aware that his investment in me is limitlessly appreciated. While I give much credit to the captain of the Ryanodine Receptor ship, the rest of the crew including Siavash Darbandi, Rebecca Vanderhooft, Rolland Gillies, Fatimah Alsaffar, and Tahani Baakdhah has been crucial in providing me with ideas and teaching me the necessary technical skills to obtain and interpret results in the laboratory. I would be amiss in not taking a moment to highlight Rebecca Vanderhooft's contribution to my project- it is through our collaborative efforts that many technical ends were achieved. I would also like to thank Dr. Wiegand for organizing this course, as well as providing me with support and feedback. His constant push to "not get to reading week without data" was vital in ensuring my preparedness to write this thesis. Thanks to Dr. S. Good-Avila and Dr. A. Civetta for their suggestions for their encouragement as members of my committee. Lastly, although by no means least, thanks to my parents, my sister, my girlfriend, and anybody else who had to suffer through long conversations about muscle contraction and ryanodine receptor expression.

Table of Contents

Abstract.....ii
Acknowledgments.....iii
Table of Contents.....iv
List of Tables.....v
List of Figures.....vi
List of Appendices.....vii
Abbreviations and Nomenclature.....viii
1. **Introduction**.....**1**
1.1 Calcium Release.....1
1.2 Structure and Function.....1
1.3 Spatial and Temporal Distribution.....4
1.4 Zebrafish.....5
1.5 Transcription Factors and Promoter Analysis.....6
1.6 Objectives.....9

2. **Materials and Methods**.....**12**
2.1 Isolating Genomic DNA.....12
2.2 Identification of RyR3 via Database Searching.....12
2.3 Polymerase Chain Reaction Amplification.....13
2.4 Cloning of PCR Products.....15
2.5 PCR Detection of Recombinants and Plasmid Purification.....15
2.6 Subcloning of Inserts into an Expression Vector System.....17
2.7 *In silico* analysis.....19

3 **Results**.....**21**
3.1 Detection of Recombinants and Sequencing.....21
3.2 MatInspector Results.....26
3.3 Match Results.....27
3.4 mVISTA Results.....29
3.5 rVISTA Results.....31

4 **Discussion**.....**36**
4.1 Bioinformatics Data.....36
4.2 Regulatory DNA Region Characterization.....39
4.3 RyR3 Promoter Comparisons.....40
4.4 Future Work.....41

5 **Conclusions**.....**42**

6 **References**.....**43**
7 **Appendices**.....**49**

List of Tables

Table 1: Primers used for amplification of 5' upstream sequence of RyR3 in zebrafish, *Danio* rerio **14**

Table 2: A) Relative position of important SNPs within an alignment of a nucleotide sequence corresponding to the upstream promoter region of RyR3, as indicated by ZV7 (Genbank; GeneID: 56150) and an assembled Contig, ZFRyR3.
B) Position of SNPs which are notably different between sequences **23**

Table 3: TF site search results using Match (gene-regulation.com)**..... 28**

Table 4: TF site search results using Match (gene-regulation.com). Core matches and matrix matches are used to minimize false positive matches.....**28**

Table 5: Transcription factor binding site hits for conserved and aligned TFs within a pairwise comparison of the RyR3 upstream promoter region of MedakaRyR3b (*O. latipes,* ENSORLG00000017871 , using a 10,000 bp upstream region of *D. rerio* RyR3 as a baseline **35**

List of Figures

Figure 1: A functional calcium release unit **2**

Figure 2: The promoter region serves the functional purpose of regulating gene expression **6**

Figure 3: Transcription factors may act as a variety of dimer types to bind DNA and regulate gene expression **7**

Figure 4: Characterization of the 5' upstream promoter region of RyR3 in zebrafish, *Danio rerio* **10**

Figure 5: RyR3 promoter region PCR product amplification visualized on a 1% agarose gel stained with ethidium bromide **21**

Figure 6: Upstream promoter sequence amplified from ZFG1, ZFG2, and ZFG4 *D. rerio* genomic DNA using primer pair DreRyR3F1 and DreRyR3R5 ..... **22**

Figure 7: PCR detection of recombinants using primer pair DreRyR3F1 and DreRyR3R5 after subcloning with a pGL4.23 expression vector **23**

Figure 8: Nucleotide sequence alignment of a region upstream of RyR3 from ZV7 (GenBank; GeneID: 56150) and a Contig, ZFRYR3 ..... **24**

Figure 9: General core promoter elements located to the RyR3 upstream promoter region using the MatInspector search engine **26**

Figure 10: Vista browser output for sequence conservation between a 10,000 bp promoter sequence of RyR3 in *D. rerio* and 8 other RyR3 promoter sequences **30**

Figure 11: rVISTA output for TFBS hits between a region 2,000-3000 bp upstream of RyR3 in *D. rerio* and *H. sapiens* **33**

Figure 12: rVISTA output for conserved and aligned TFBS hits between a region 2,000-3000 bp upstream of RyR3 in *D. rerio* and *O. latipes* RyR3b **34**

List of Appendices

Appendix A: 1Kb DNA ladder.....**49**
Appendix B: pGEM-T Easy Vector System Map.....**50**
Appendix C: pGL4.23 Expression Vector System Map **52**
Appendix D: Restriction Enzyme (RE cut sites) found within ZFRYR3 by NEBcutter2.0, along with compatibilities of a selection of enzymes with pGL4.23 .....**54**
Appendix E: Detailed BLASTn search results.....**56**

Abbreviations and Nomenclature

AA: Amino Acid

bHLH: Basic helix-loop-helix motif, including a wide variety of proteins such as myoD, MEF-2, myogenin, usf, mycmax, and HAND.

bp: Basepair, a unit of measurement used for DNA length

cis-regulatory: regions of DNA that regulate expression of genes, and generally contain binding sites for one or more transcription factors (TF).

Ca^{2+}: Calcium ions

Mg^{2+}: Magnesium ions

CAAT box: A DNA sequence located approximately 75 base-pairs upstream from eukaryotic transcription start sites. This sequence is one of those that enhance binding of RNA polymerase

CART: Cocaine-amphetamine regulated transcript

CLOX: Cut-like homeobox TF

Contig: from the word "contiguous"; a series of overlapping DNA sequences used to make a physical map that reconstructs the original DNA sequence of a chromosome or a region of a chromosome. A contig can also refer to one of the DNA sequences used in making such a map.

DHPR: Dihydropyridine receptor

DNA- Deoxyribonucleic acid

E-box: Sequence bound by basic helix-loop-helix transcription factors; a canonical sequence of 5'-CANNTG-3'

E-C coupling: Excitation-contraction coupling

EVI: Ectopic viral integration protein

FOX: Forkhead box domain

GATA: TF family capable of binding the sequence GATA

hpf: Hours post-fertilization

Abbreviations and Nomenclature (continued)

HNF: Hepatocyte nuclear factor

IP_3R: Inositol-1, 4, 5-triphosphate receptor

Ladder (1 kb): A DNA standard used for size comparison on an agarose gel stained with ethidium bromide; an appropriate ladder will contain digested fragments within an appropriate size range for isolating DNA fragments of interest

mRNA: Messenger ribonucleic acid

MCS: Multiple cloning site

NKX: Natural Killer X; A homeodomain TF

NRF: Nuclear respiratory factor

OCT: Octameric-binding protein

PCR- Polymerase chain reaction

RE: Restriction Enzyme

RET: Receptor tyrosine kinase

RyR: Ryanodine Receptor

S.N.A.P.: Gel purification kit produced by Invitrogen

STAT: Signal transducer and activator of transcription factor

SP1: A zinc-finger motif protein which interacts with bHLH proteins

TATA box: A core promoter element found in the promoter region of genes in eukaryotes

TF: Transcription factor

TFBD(S): Transcription-factor binding domain (site)

Upstream: 5' with respect to a given sequence of DNA

1. Introduction

1.1 Calcium Release

Calcium is an important second messenger involved in a plethora of developmental and fundamental processes including fertilization, synaptic transmission, and muscle contraction (Ashworth *et al,* 2001; Ashworth 2004). Two distinct types of calcium release channels have been found in vertebrates; the Inositol-1,4,5-triphosphate receptor (IP3R) and the ryanodine receptor (RyR; Ashworth *et al.,* 2007; Fill and Copello 2002). Each of these proteins is named for a particular ligand the receptor is known to bind. In the process of muscle contraction, a signal transduction pathway is propagated by an electrochemical signal which acts at the neuromuscular junction (Koulen *et al,* 2001). This causes depolarization along the myocyte cell membrane, initiating the process of mechanically-induced calcium release or calcium-induced calcium release from the sarcoplasmic reticulum (SR) into the cytoplasm, increasing intracellular [Ca^{2+}] concentration (Fill and Copello 2002). This process is known as excitation-contraction (EC) coupling (Fill and Copello 2002).

1.2 Structure and Function

Ryanodine receptors, named for the plant alkaloid ryanodine which binds them, are large homotetrameric receptor proteins found to have a larger conductance of calcium than IP3Rs (Giannini *et al,* 1995). The RyRs are transmembrane proteins localized to the sarcoplasmic reticulum of myocytes and the endoplasmic reticulum (ER) of other cell types including neurons (Thibault *et al,* 2007; Stutzmann *et al,* 2006). Each RyR is approximately 2.2 MDa in size (Giannini *et al,* 1995). Closer analysis shows that

these RyRs are usually localized to so-called “triad junctions”, consisting of RyR1, RyR3 and the voltage-gated Dihydropyridine receptor (DHPR; Perez *et al,* 2005). RyR binding to each DHPR involves a specialized binding region or “foot” structure (Paolini *et al,* 2004; Figure 1). They may be ligand-gated, as is seen in cardiac muscle, or opened by mechanical/depolarization-induced activation, as has been described in skeletal muscle (Protasi 2002). This mechanical activation, while previously referred to as mechanically-induced calcium release, is commonly also referred to as depolarization-induced calcium release (DICR; Koulen *et al,* 2001).

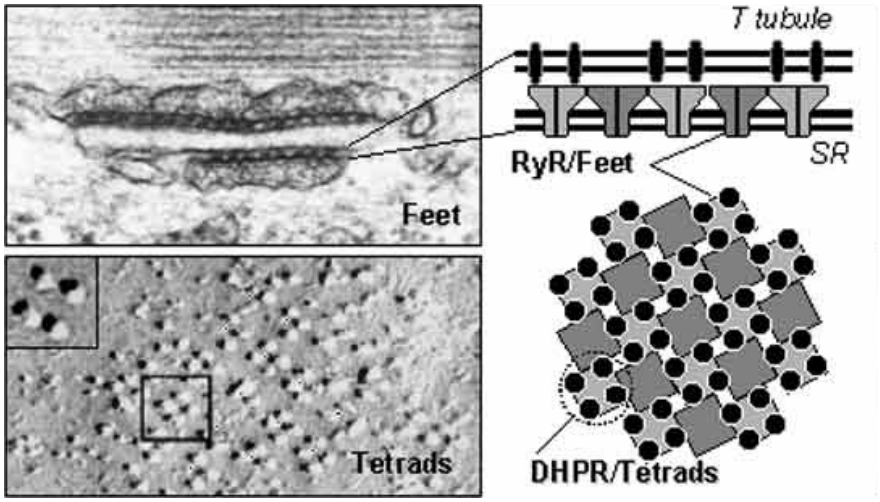


Figure 1: A functional calcium release unit. Shown on the right is the association of the RyR foot domain with DHPR binding in tetrads. Left: An electron micrograph showing the foot structure and tetrads of RyR1, RyR3 and DHPR. Adapted from Protasi 2002.

It is known that RyR channels share common functional properties, including regulation by channel modulators such as Ca^{2+}, Mg^{2+}, and caffeine (Ashworth and Bolsover 2001). Furthermore, it has been shown that physiologically different RyR forms can be found localized to distinct muscle tissue types, as well as distinct cell types in non-muscle cells (Mattei *et al,* 2004). This tissue distribution, concurrent with different modes of action, corresponds to different isoforms of the RyR; in mammalian

vertebrates, these are designated RyR1, RyR2, and RyR3. RyR1 is predominantly expressed in skeletal muscle, while RyR2 is expressed in cardiac muscle, smooth muscle, and certain areas of the brain (O'Brien *et al.,* 1993). Of prime interest is RyR3, a ubiquitously-expressed isoform which is expressed significantly in developing mammals, followed by downregulation in adults (Giannini *et al,* 1995). Work done by Perez and colleagues (2005) showed that in cultured myotubes and muscle fibers, expression of RyR3 can be associated with an augmented spontaneous Ca^{2+} activity, as well as with increased frequency and size of Ca^{2+} sparks, the Ca^{2+} release events that occur within muscle cells. They also found that caffeine sensitivity and increased calcium concentration are not the result of variations in expression levels of channels but arise from intrinsically distinct properties that are characteristic of each isoform. Thus, it is important to analyze each isoform individually, in order to understand its properties more precisely.

Non-mammalian vertebrates such as zebrafish (*Danio rerio)* are of interest as they add another dimension to the expression biology of the RyRs. The RyRs in fish are localized to specific muscle fiber types; slow-twitch red muscle (RyR1a) and fast-twitch white muscle (RyR1b; Franck *et al,* 1998; Nishida *et al,* 1996). Early efforts to characterize expression of RyRs in fish found two isoforms expressed together in skeletal muscle- an α-isoform physiologically similar to RyR1, and a β-isoform physiologically similar to RyR3 (O'Brien *et al,* 1995). As in mammals, RyR3 expression is observable in a variety of zebrafish tissues, including portions of the brain and skeletal muscle (Fill and Copello 2002).

1.3 Spatial and Temporal Distribution of RyR3

It is known that RyR3 is coexpressed with RyR1 in fast-twitch muscle in mammals (Nabhani *et al,* 2002). Studies on murine specimens have shown that RyR3 mRNA can be detected in a variety of tissue types, including the cerebrum, cerebellum, lung, esophagus, spleen, gut, kidney, stomach, thymus, adrenal gland, and ovary (Giannini *et al,* 1995). RyR3 has also been found to participate in differentiation of myoblasts to skeletal muscle cells, suggesting that this particular isoform may play a significant role in development (Nabhani *et al,* 2002). In fact, the maintenance of the skeletal muscle lineage requires RyR3 expression, and absence of RyR3 expression prevents differentiation from myoblasts (Nabhani *et al,* 2002).

Physiological properties of RyR3 have also been described via immunoblot analyses and single channel recordings of SR preparations (Koulen *et al,* 2001). Studies are needed to better understand temporal and spatial distribution of RyR3 expression in non-mammalian vertebrates. While transcriptional level control of some RyR genes has been described in other animals, such as human RyR1 and rabbit RyR1, such control mechanisms have not been described to date in the regulation of RyR3 in fish (Zorzato *et al,* 1990). Certain studies have pursued information in this regard; for example, it is known that RyR3 expression is not detectable in the gastrula stage, but is expressed in myotomes 24-48 hours post-fertilization (hpf; Ward *et al.,* 2001). Additionally, it is known that pharmacological blocking of the RyR by ryanodine binding will affect myofibril organization (Brennan *et al,* 2005). A recent study performed by Darbandi and Franck found that RyR1b and RyR3 are co-expressed at equivalent levels in certain

Zebrafish tissues (2009). Work must be done to further understand how RyR3 expression is controlled in a variety of tissue types, and what these control mechanisms may imply about RyR3 function.

1.4 Zebrafish

Zebrafish is an organism often used in molecular genetics due to its small size and rapid generation time (He *et al.,* 2006). Furthermore, the existence of the zebrafish genome project allows for easy and rapid database searching. Thus, it is logical to focus on the identification of *cis*-acting regulatory domains for RyR3 in *D. rerio.* These *cis*-regulatory elements are regions of DNA that regulate expression of genes, and generally contain binding sites for one or more transcription factors (TF). The RyR3 gene in *D. rerio* has been identified from the **Ensembl** zebrafish genomic database (OTTDARG00000005601) and is mapped to Chromosome 20 p33,626,994 – 33,756,627. Expression analysis is also simplified through the use of *D. rerio,* as several types of cell lines are available which have been produced from different types of *D. rerio* cells, including both a fibroblast-like cell line (ZF4) and a liver cell line (ZFL), among others (He *et al.,* 2006). These cell lines are readily available through the American Type Culture Centre (ATCC) and are relatively untransformed compared to the cells they were derived from. Of further interest, *D. rerio* embryos are transparent, allowing for the easy visualization of microinjected embryos 24-48 hpf (Stuart *et al.,* 1990). This provides a major advantage, since it may be possible to track localization and relative amount of RyR3 during development.

1.5 Transcription Factors and Promoter Analysis

In order for transcription to take place, the enzyme that synthesizes RNA, known as RNA polymerase, must associate with the DNA upstream of a gene (Figure 2). Promoters generally contain specific DNA sequences and response elements which provide a binding site for RNA polymerase as well as TF binding sites (Karp 2008). These TFs are diffusible factors which are usually proteins that may modify the expression of genes distant from the gene that was originally transcribed to create them (Stem 2000).

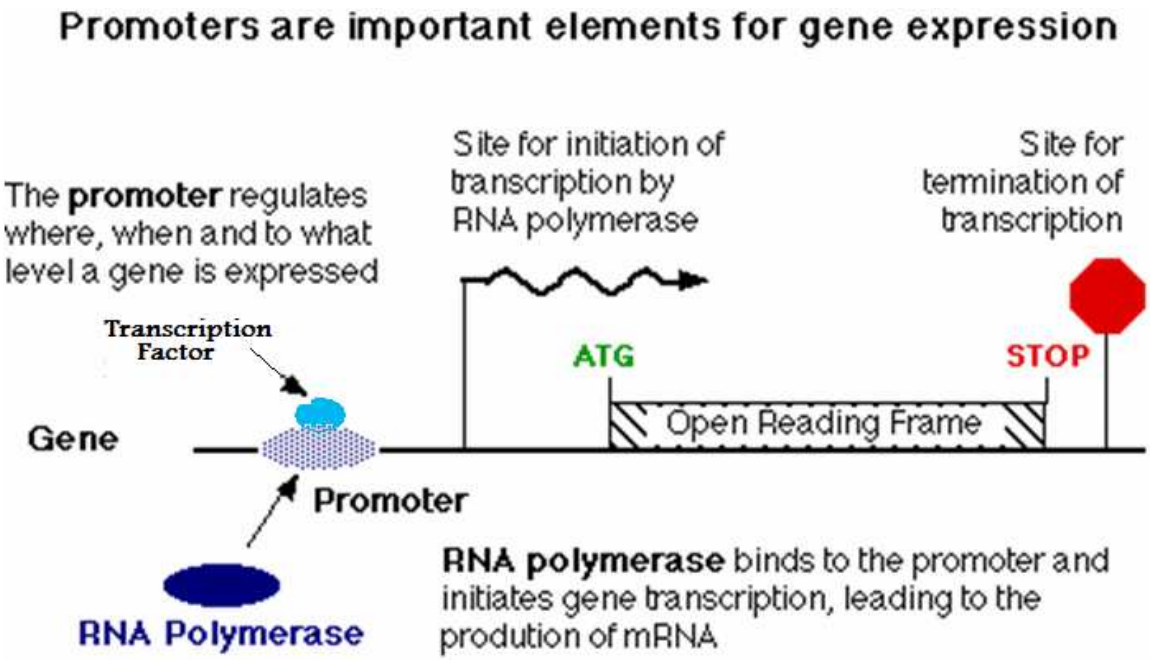


Figure 2: The promoter region serves the functional purpose of regulating gene expression. Shown above is the relationship of the promoter to the initiation site, as well as an indication that the binding site for RNA Polymerase may be dictated by certain promoter elements, including the TATA box. Adapted from Karp 2008.

TFs are known to interact with each other in cascade events as well, increasing the complexity of any regulatory system (Dermitzakis and Clark 2002). They may be grouped into categories of greater specificity, depending on function. The basic helix-loop-helix (bHLH) TF group includes a host of proteins such as myogenin, myoD, myf5,

herculinin, and MEF-2 (Edmondson *et al,* 1992). While TFs in this group bind to the DNA consensus sequence CANNTG (also called an E-box), individual HLH proteins may act as homodimers, displaying half-site preferences that depend on nucleotides within and surrounding the consensus sequence, a genetic sequence found with minor variations and similar functions in widely divergent organisms or genetic locations (Vavouri and Elgar 2005; Figure 3). Since RyRs are expressed in muscle, binding sites for these TFs may be found in the promoter sequence, as these TFs are often involved in cross regulation of muscle-specific genes (Edmonson *et al,* 1992).

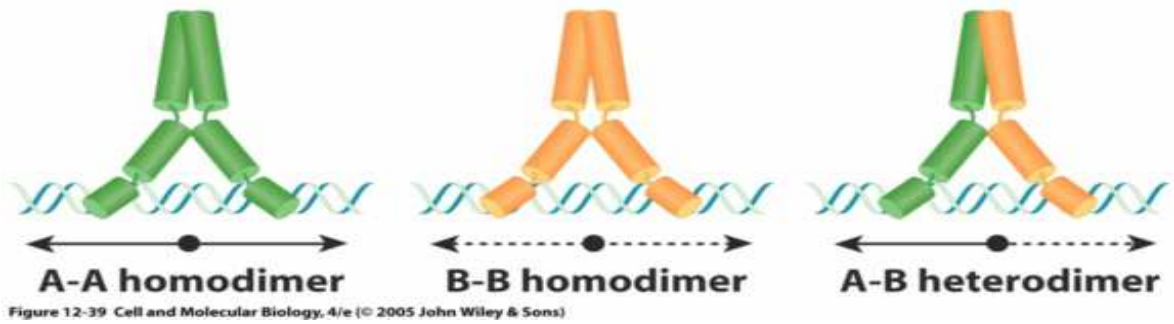


Figure 3: Transcription factors may act as a variety of dimer types to bind DNA and regulate gene expression. Proteins of the basic helix-loop-helix (bHLH) may bind and interact with one another as homodimers in the case of muscle-specific transcription factors, or as heterodimers in certain cases of development. The basic region of the bHLH type TFs may be absent, causing binding to a basic-domain containing TF to nullify function entirely, as is observed in differentiated neural progenitors. Adapted from Karp 2008.

While a variety of TF groups exist, one particularly important group is called the forkhead box (FOX) family of TFs (Lehmann *et al,* 2003). This group of TFs binds to the following sequence: 5'-A[AT]T[AG]TTTGTTT-3' and each may act as either a transcriptional activator or repressor (Lehmann *et al,* 2003). Previous studies have shown that these TFs, including FOXF2, FOXD1-6, and FOXO all initiate and maintain development of neural crest cells from neural tube progenitors in early development

(Brunet *et al,* 2004). Other work shows that FOX transcription factors are absolutely required to maintain pluripotent stem cells in both pre- and peri-implantation states in embryogenesis (Lehmann *et al,* 2003). Sites for FOX TF binding may be expected in a functional promoter region for RyR3, given that it may play a role in the development of cells in contractile and non-contractile tissues. Work to identify sites in the promoter region of RyR3 in *D. rerio* for any of the aforementioned TFs has not been performed to date.

The role of promoters as critical elements that can work in concert with other regulatory regions including enhancers, silencers, and insulators is also of critical interest (Drivenes *et al.,* 2000). These interactions may be involved in upregulating or downregulating the level of transcription of a given gene (Karp 2008). Within such a promoter region, certain elements may be found including the TATA box, a core DNA sequence 5'-TATAAA-3' (or a variant) usually located 25 base pairs upstream of the transcription site (Sawada *et al,* 2008). Similarly, the CAAT box, a DNA sequence located approximately 75 base-pairs upstream from eukaryotic transcription start sites that enhances binding of RNA polymerase, may also be located in this region (Dermitzakis and Clark 2002).

It has been shown by previous studies of mammalian RyR1 and RyR2 that *cis*-regulatory domains are located within 1 kilobase (kb) of the transcription start codon (Schmoelz *et al,* 1996; Nishida *et al,* 1996). Other muscle-specific genes such as myogenin have promoter elements within a relatively short stretch of approximately 1 kb 5’ to the initiation site (Edmondson *et al,* 1992). It is known that many muscle-

specific genes do not contain a CAAT box or TATA box 30 bp upstream of the initiation site (Edmondson *et al,* 1992). Additionally, binding sites for bHLH transcription factors such as myoD or myf5 are usually found in the promoter regions of muscle-specific genes (Davis *et al,* 1990).

In order to search for TFBDs in a region upstream of the RyR3 gene in zebrafish, bioinformatics programs were used. These programs, MatInspector (Genomatix), Match (gene-regulation.com), and the VISTA suite (Dubchak and Ryabov 2006), allow for the identification of potential binding sites in the upstream regions of coding sequences using database information compiled from other studies. Many of the aforementioned transcription factors binding sites may be found in the 5' upstream region of RyR3 (Divine *et al,* 2003; Wieser 2007; Lee *et al,* 2001).

In order to study the promoter's ability to drive transcription, the sequence can be ligated into a promoterless expression vector. These vectors, such as pGL4.23, contain a gene for luciferase protein which can be visualized and quantified, where the quantity of luciferase expression reflects the promoter's ability to drive transcription. These vectors may be transfected into cell lines to study spatiotemporal expression. More importantly, these expression vectors can be stored for use in functional studies proceeding this work and serve as confirmation tools for any inferences made from computer-based analyses.

1.6 Objectives

The main objective of my research was to characterize a 2 kb region of sequence immediately upstream of the first exon in *D. rerio* RyR3. This sequence was identified via

database searching, amplified from *D. rerio* genomic DNA samples using appropriate primers, and subcloned into an expression vector for further expression studies. Using computer based (*In silico)* analysis, 5' upstream sequences for RyR3 gene were aligned between a variety of mammalian and non-mammalian vertebrate species, to determine whether any sequence conservation is present. Further to this, putative transcription factor binding sites were characterized through the use of the VISTA genome browser and two other databases; namely, MatInspector (Genomatix) and Match (Waleev *et al,* 2006; Figure 4).

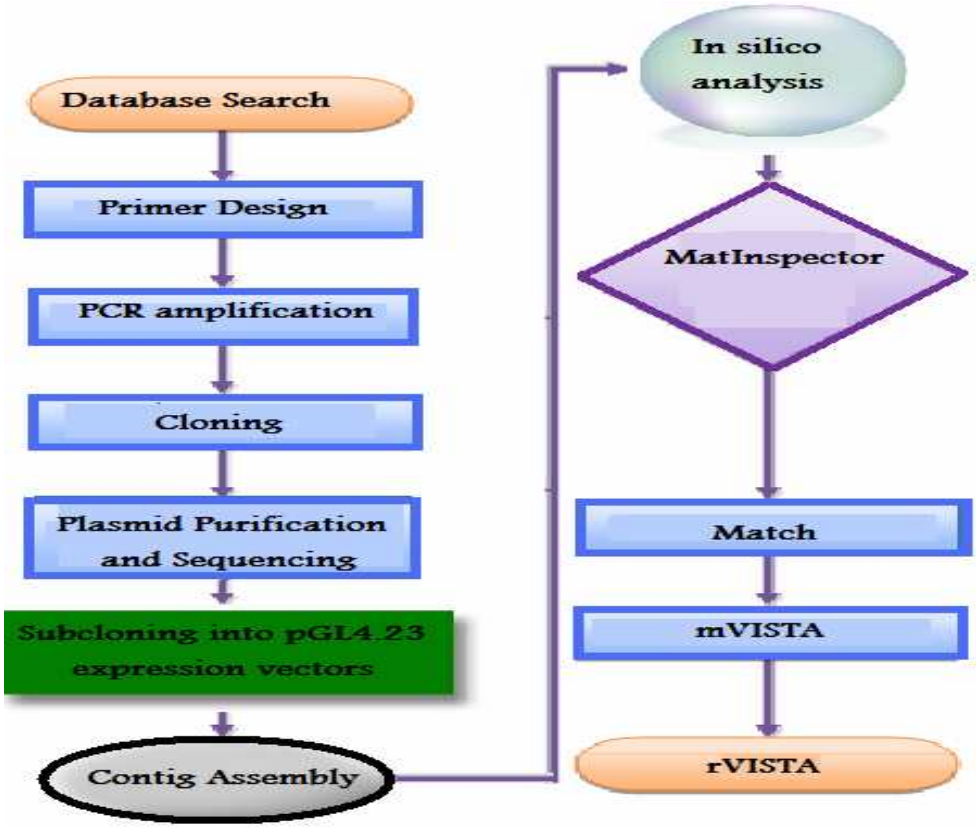


Figure 4: Characterization of the 5' upstream promoter region of RyR3 in zebrafish, *Danio rerio.* Shown above are the steps which were taken in obtaining stored clones of vectors containing RyR3 promoter, as well as steps in computer based (*In silico* analysis). Further clarification of details pertaining to each step is included in Materials and Methods.

It was expected that nucleotide regions which were predicted to be extremely important in the regulation of muscle-specific genes, such as E-boxes, should be found in abundance within this promoter region. Sites bound by TFs related to developmental regulation were also surveyed in this study to further clarify RyR3's implicit role in early growth processes. By surveying for such sites in conserved regions, it is possible to understand which TFBDs are crucial for either transcriptional activation or repression.

The characterization of these *cis*-regulatory regions in RyR3 will aid in more clearly outlining the complex relationship between RyR3 and other elements involved in development, providing important genetic insight into the expression of this component of vertebrate EC coupling.

2. Materials and Methods

2.1 Isolating Genomic DNA

Genomic DNA was isolated from frozen whole zebrafish preparations, obtained from a local tropical fish store. Fish specimens were sacrificed using a protocol approved by the Senate Animal Care Committee issued to Dr. Franck (April 2009). Tissue samples of 0.5 mg were mixed with 1 mL of DNAzol to isolate genomic DNA. Additionally, Proteinase K (100 µg/mL) was also added in order to digest cellular components which may interfere with isolation of genomic DNA. After digestion overnight at 25^0C, differential centrifugation was used to obtain DNA only, including a spin at 10,000 x g at 25^0C for 10 minutes followed by DNA precipitation using 0.5 mL 95% ethanol/1 mL DNAzol used. DNA was then pelleted by centrifugation using one spin at 5000 x g at 25^0C; after this step, the supernatant was removed. DNA was washed twice with 1 mL 75% ethanol, followed by a dissolving step in 0.3 mL 8 mM sodium hydroxide. HEPES was also added to isolated genomic DNA samples to adjust pH to 8. These samples were prepared in March 2009, and labeled as ZF genomic 1, 2, 3, and 4 (ZFG1, 2, 3, and 4). Of these, ZFG1, ZFG2, and ZFG4 were used.

2.2 Identification of RyR3 via Database Searching

A preliminary search was performed using Genbank (www.ncbi.nlm.nih.gov) using the search terms “Zebrafish RyR3.” In total, 8 entries were found, including the gene itself. The gene may be found as GeneID: 561350 with a location on Chromosome 20, reverse strand. Total length was shown to be 150,857 bp with a processed length of

14,592 bp. The corresponding protein was shown to be 4,863 AA long after processing. The first eight exons of this protein were compared to the same exons found in RyR3 in a variety of other fish species, including medaka, tetraodon, and fugu, as well as mouse and human, in order to ensure the protein was RyR3 (not included). Approximately 2000 bp of the 5' upstream sequence was taken, along with approximately 200 bp into the first exon, including the presumed translation start site (a site preceding the first exon including the nucleotide sequence ATG). Total sequence length was 2367 bp including the 200 bp portion of exon 1. An entry for zebrafish RyR3 on Ensembl (Ensembl entry: ENSDARG00000042531) was also identified and compared to ensure no differences between major databases in this region; the sequence was found to be nearly identical (not included).

2.3 Polymerase Chain Reaction Amplification

Primer pairs were designed using Primer3 (http://frodo.wi.mit.edu/primer3/) to amplify fragments of DNA of approximately 500 bp in length of the previously described 2000 bp region upstream of RyR3 in Zebrafish. Primers were ordered from Invitrogen (Table 1).

PCR was performed in 600 µL micro-centrifuge tubes, each of which contained the following reagents partitioned from a master mix: 2.5 µL 10X PCR buffer, 0.75 µL 50 mM $MgCl_2$, 0.5 µl 10 mM dNTPs, 18 µl ddH_20 (RNase free water), and 0.25 µl (1 U) *Taq* DNA polymerase (Invitrogen Life Technologies) for a total volume of 22 µL per reaction. To each reaction, 0.5 µL 100 µM forward and 100 µM reverse primer along with 3 µL 20 ng/µL template were added. Each reaction tube was then gently mixed and briefly

centrifuged at 2000 x g at 25^0C. Tubes were then inserted into a thermocycler system. The program used was 35 cycles long, including the following 3 steps: denaturation for 1 minute at 95^0C, annealing for 1 minute 30 seconds at 55^0C, and extension for 1 minute 30 seconds at 72^0C. An initial unrepeated step involving 9 minutes of denaturation at 95^0C and a 7 minute extension step at 72^0C were also included in the PCR. All reactions were held at 4^0C indefinitely after completion of the additional extension step. After this, 8 µL of experimental reactions and 8 µL of 1 Kb DNA ladder (Appendix A) were loaded onto a 1% agarose gel stained with ethidium bromide, and visualized under UV light using a GelDoc-RX camera (BioRad). This step is necessary in determining the success of amplifying the 2 kb upstream region, as well as in determining whether any contamination has occurred. PCR products were then purified using an S.N.A.P. gel purification kit (Invitrogen Life Technologies).

Table 1: Primers used for amplification of 5' upstream sequence of RyR3 in zebrafish, *Danio rerio

Name	Primer Sequence (5'-3')
DreRyR3F1[a]	CTGGTTAGGGGATTTGTAGC
DreRyR3R1	TCAATTTAGGGCAGGATTTG
DreRyR3F2	TACCACCTTGCAGAAGCAC
DreRyR3R2	CACTGTGGTGGGATCTTTTC
DreRyR3F3	GCAAATCCTGCCCTAAATTG
DreRyR3R3	TAGGTGAAACTGGTCCTTGG
DreRyR3F4	TACCAAGGACCAGTTTCACC
DreRyR3R4	GGTGCTTCATTTGCTTCATC
DreRyR3F5	GTATTGGGACAGCATGAAGG
DreRyR3R5[a]	AGCGGCGTTTTTACTTACAG
REDreRyR3F1[b]	**CGCGGTACC**CTGGTTAGGGGATTTGTAGC
REDreRyR3R5[b]	**CGCCTCGAG**AGCGGCGTTTTTACTTACAG

*-The letter "F" in the primer name indicates a forward primer and an "R" indicates a reverse primer.
a- DreRyR3F1 was used with DreRyR3R5 to obtain approximately 2 kb fragment for cloning into pGEM-T easy vector.
b- Primer set designed to add restriction enzyme sites to 2 kb fragment, to ligate into pGL4.23 expression vector. Recognition sites have been included in each primer for KpnI (ReDreRyR3F1) and XhoI (ReDreRyR3R5) respectively, along with an added triplet, CGC. Bolded nucleotides contain the RE recognition site.

2.4 Cloning of PCR Products

Purified PCR products were ligated into the plasmid vector pGEM-T easy from Promega Biotech, and then transformed into *Escherichia coli* JM109 competent cells (Appendix A). Cells were then plated onto NCZYM media supplemented with 0.1 M isopropyl β-D-1 thiogalactopyranoside (IPTG), 40 mg/mL X-gal (a lactose analogue), and 50 mg/mL ampicillin. The IPTG acts as an inducer of the *lac* operon in *E. coli* while X-gal is used to visualize β-galactosidase activity. Those colonies which contained only uncut plasmid appear as blue, since the *lacZ* gene is uninterrupted in such plasmids; white colonies are selected for purification and sequencing therefore, as these colonies would contain a plasmid with the promoter region cloned into the MCS, disrupting the *lacZ* gene. Additionally, ampicillin is also added to such media since JM109 cells are ampicillin-sensitive, but plasmids contain an antibiotic resistance element, promoting the growth of cells that have undergone successful transformation with the pGEM-T easy vector system (Appendix B). Inoculation was performed using aseptic technique, followed by incubation for 18-24 hours at 37^0C.

2.5 PCR Detection of Recombinants and Plasmid Purification

A PCR-detection mixture is used to ensure that the expected size inserts cloned into the MCS of pGEM-T easy. To assess this transformation, white colonies were isolated and assessed for insert size to ensure selection excluded small satellite colonies. The following mixture was used for this PCR detection: 1.5 µL 10x PCR buffer, 0.9 µL 50 mM $MgCl_2$, 0.075 µL 100 µM M13 forward primer, 0.075 µL 100 µM M13 reverse primer, 119.25 µL ddH_20 and 1.5 µL (1 U) *Taq* DNA polymerase. M13 primers were

designed to amplify the insert region of the plasmid (Appendix B). A portion of the plasmid was also amplified along with each insert due to primer location, therefore leading to a slight increase in the size of the expected products (Figure 5). PCR detection was performed using 35 cycles, and was composed of the following three main steps: 1 minute at 95^0C, 30 seconds at 55^0C and 1 minute at 72^0C. An additional 5 minute denaturing step at 95^0C was also included at the beginning of the program along with a 7 minute extension step at 72^0C to cap the process. The PCR products were then visualized on 1% agarose gel stained with ethidium bromide and viewed under UV light using the GelDoc apparatus, as was performed after initial amplification from genomic DNA. Colonies found to have the correct insert size, approximately 500 bp, were grown in a shaker overnight at 37^0C in an NCZYM broth supplemented with 0.05 mg/mL ampicillin (Figure 5). Isolation of plasmids was then performed using the PureLink Quick Plasmid Miniprep Kit following manufacturer protocol (Invitrogen Life Technologies). Each plasmid preparation was quantified in ng/µl using a nanospectrophotometer to ensure equivalent amounts of plasmid DNA within each tube (35 ng/µL). Sequencing of the plasmids was performed by The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada (TCAG).

Cloning of PCR Products was performed with minor changes to obtain a single fragment of approximately 2 kb in size (Figure 6). In this case, PCR amplification differed in several respects. The primer pair DreRyR3F1 and DreRyR3R5 was used exclusively and PCR reactions were performed with only 0.4 µL of 20 ng/µL template. The PCR master mix components were also different, as attempts to amplify 1.8 kb with DreRYR3F1 and

DreRyR3R5 using *Taq* DNA polymerase were initially unsuccessful. Thus, a different DNA polymerase with higher fidelity and added proofreading activity called Platinum *Pfx* DNA polymerase was used instead. PCR reactions were still performed with the following mixture: 1 µL 50 mM $MgSO_4$, 1.5 µL 10 mM dNTPs, 40.2 µL ddH_20, 5 µL 10x *Pfx* buffer, and 0.4 µL 2.5 U/µL Platinum *Pfx* DNA polymerase. To each tube, 0.75 µL of 100 µM primer was added along with the insert template DNA. Thermocycler settings were as follows: denaturation for 15 seconds at 94^0C, annealing for 30 seconds at 56^0C, and extension for 2 minutes at 68^0C. These three main steps were cycled 35 times, with an initial 94^0C denaturation step lasting 5 minutes, and a 7 minute extension at 68^0C following cycling. Temperature was set to hold at 4^0C indefinitely after completion. After gel purification was performed with a S.N.A.P. gel purification kit, DNA was subjected to an additional necessary step to add a 3' Adenine (A) overhang, in order for ligation to work properly as per instruction from Invitrogen. To each sample, 2.5 µL 10x PCR buffer, 0.5 µL 10 mM dATP, and 0.25 µL (1.25 U) *Taq* DNA polymerase were added, followed by a 72^0C extension step for 15 minutes. Ligation, detection of recombinants, and plasmid purification were performed as previously indicated, and were sequenced after quantification at TCAG.

2.6 Subcloning of Inserts into an Expression Vector System

Expression vectors were obtained from Dr. Marc Ekker at the University of Ottawa, including pGL4.23 (Appendix C). The MCS of pGEM-T easy and the MCS of pGL4.23 do not contain any analogous restriction enzyme (RE) sites. All restriction enzymes for pGL4.23 were compared to any RE sites found in the sequence of RyR3,

using a program called NEBcutter2.0 (New England Biolabs; Appendix D). Two enzymes, KpnI and XhoI, were found in the MCS of pGL4.23, and not found within the RyR3 promoter sequence. In order to subclone from pGEM-T easy, primers were designed to include a site for KpnI and a site for XhoI, as well as 3 extra bases (CGC) in order to ensure product stability. PCR amplification was performed as previously described from pGEM-T easy with REDreRyR3F1 and REDreRyR3R5 as primers, with the exception of a 58^0C annealing step in PCR. Double digests were then performed in 600 µL microcentrifuge tubes, with the following ingredients partitioned from a master mix: 2.5 µlL A buffer (Roche diagnostics), 0.25 µL 10 mg/mL Bovine Serum Albumin (BSA), 3.25 µL ddH_20, and 18 µL 30 ng/µL insert DNA.; 0.5 µL 10 U/µL XhoI and 0.5 µL 10 U/µL KpnI were added last. To digest the plasmids, BSA, buffer A, and enzymes were added in the same quantities, while 15 µL 20 ng/µL pGL4.23 plasmid and 6.75 µL ddH_20 were used. Digests were run using a PCR program involving incubation at 37^0C for 2 hours, followed by a 15 minute heat inactivation step at 65^0C. Double digested plasmid was purified from a LMP gel. After heat inactivation, the 25 µl digests were removed from the thermocycler, immediately placed in chilled water, and ligation reactions were performed in the same tube. Ligation runs were performed, however, with the addition of 2.5 µL 10 mM dTT and 0.6 µL 100 mM ATP, followed finally by 1.5 µL 3 U/µL T4 DNA ligase. Approximately 2 µL 200 ng/µL pre-digested plasmid was added to this mixture. Transformations were performed as previously described. Identification of *E. coli* JM109 colonies transformed with uncut plasmid versus those transformed with an insert-containing vector is made extremely difficult due to the lack of a specific marker. Thus in

detecting recombinants, 14 different colonies were screened with PCR primers to survey for a 2 kb insert (Figure 7). Detection of recombinants was followed by plasmid purification, nanospectrophotometer quantification, and sequencing at TCAG. All permanent clones were stored at -80^{0}C in glycerol stocks; plasmid preparations were stored at -20^{0}C.

2.7 *In silico* Analysis

After sequencing, the 2 kilobase region was analyzed for location and abundance of all available vertebrate transcription factors using the MatInspector database. This software tool uses a vast library of matrix descriptions for TFBDs to locate matches in DNA sequences. MatInspector assigns a quality rating to each match, thus allowing for quality-based filtering and selection of matches. The program is able to do so based upon its grouping of transcription factor binding sites into matrix families based on functional similarity. Mismatches may be avoided by toggling the parameter “core similarity”, a term used to describe a TFBDs sequence identity to a precomputed, highly conserved four base pair sequence in the anchored center of a given matrix. A decrease in core similarity allows for a lower degree of specificity for the core region of a specific TF, while an increase in core similarity causes the program to overlook potential matches. Additional features of this software include the inclusion of other lines of evidence for putative TFBDs. After the application of MatInspector, a program called Match was used (Waleev *et al.,* 2006). Match is a program which makes use of the TransFac database, a compilation of putative binding sites of known muscle-specific and developmentally related transcription factors. This Transfac database is publically

available and includes the most comprehensive cross-species compilation of TFs published to date, curated using experimentally derived data about each TF and its properties. All position weight matrices set up based upon the TransFac database are therefore regarded to be the most accurate for bioinformatics research (Waleev *et al,* 2006).

Finally, upstream sequences were located by database searching from a more extensive 10,000 bp region upstream of the RyR3 gene of *D. rerio* and several other mammalian and non-mammalian vertebrate species, including fugu (*Takifugu rubripes*), tetraodon (*Tetraodon nigroviridis*), medaka (*Oryzias latipes*), and human (*Homo sapiens*). These sequences were analyzed together using a program called main VISTA (mVISTA) to identify areas of sequence conservation between promoters. Regulatory VISTA (rVISTA) was utilized to survey for transcription factor binding site hits in any of the nine compared sequences (Dubchak and Ryabov 2006). Searching for TF sites using regulatory VISTA was performed using TF sites identified by Match searching in ZFRyR3 This program uses the Transfac database, as previously described, but is also capable of scanning for TF sites in conserved regions between promoters. VISTA allows for the comparison of up to 100 promoter sequences from a variety of sources, including uploaded sequences and precomputated genomes available on their server (Dubchak and Ryabov 2006). Using a threshold cutoff of 85% core similarity from MatInspector results, these RyR3 promoter regions were surveyed for a selection of 30 different TF sites.

3. Results

3.1 Detection of Recombinants and Sequencing

From the isolated genomic DNA sample ZFG2, PCR products were successfully amplified using primer set DreRyR3F1 through DreRyR3R5 to obtain overlapping fragments of approximately 500 bp in length (Figure 5). At least one lane produced from each set of primer pairs showed successful amplification, although certain lanes contain bands at distinctly different size regions due to mispriming and primer dimerization (Figure 5). Additionally, using DreRyR3F1 and DreRyR3R5, a fragment of approximately 2 kb was successfully isolated using genomic DNA samples ZFG1, ZFG2, and ZFG4 (Figure 6). This sequence is stored in permanent clones containing pGEM-T easy vectors for future subcloning of promoters into vectors of choice.

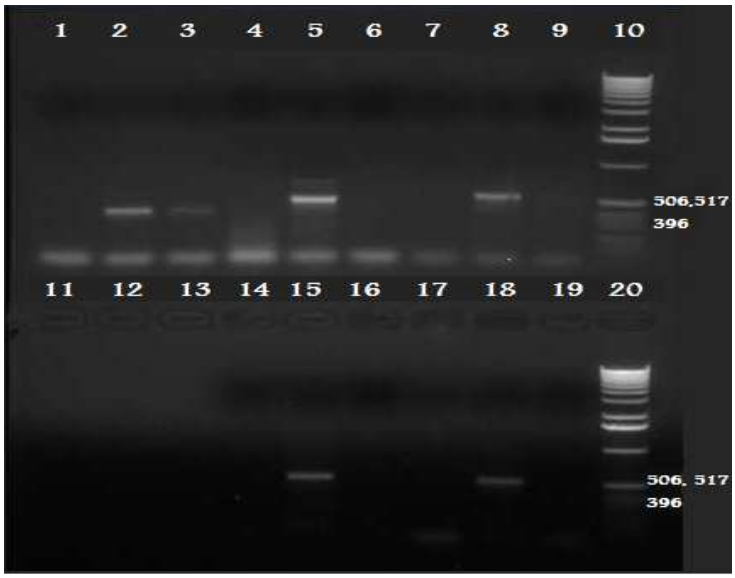


Figure 5: RyR3 promoter region PCR product amplification visualized on a 1% agarose gel stained with ethidium bromide. Each primer pair was tested three times for confirmation, in adjacent lanes. Lanes contain sequences amplified with the following primer sets, each beginning with the notation DreRyR3X:F1-R1 (Lane 8), F2-R2 (Lane 5), F3-R3 (Lane 2, 3), F4-R4 (Lane 18), F5-R5 (Lane 15). Lanes 10 and 20 contain 1 kb ladder loaded for visual estimation of band sizes. Relevant fragment sizes on reference ladder are indicated in white to the right of each ladder in base pairs (bp).

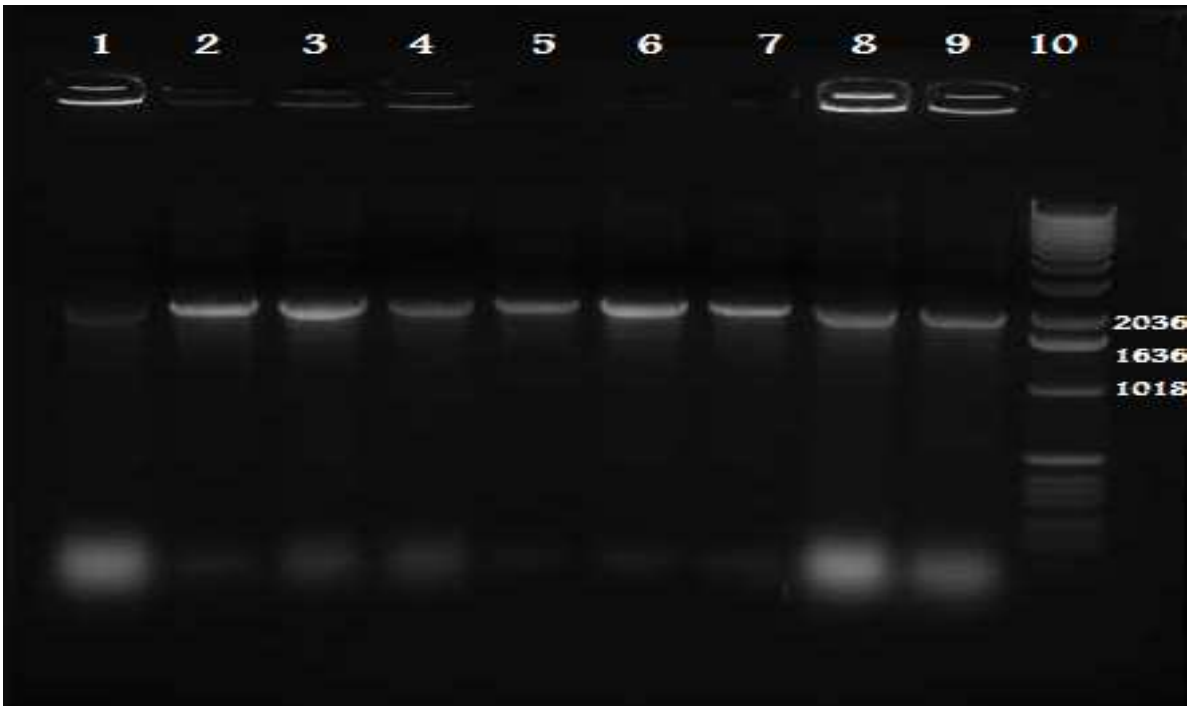


Figure 6: Upstream promoter sequence amplified from ZFG1, ZFG2, and ZFG4 *D. rerio* genomic DNA using primer pair DreRyR3F1 and DreRyR3R5. PCR product amplification was visualized on a 1% agarose gel stained with ethidium bromide. All products appear to be consistently within the size range of 2 kb, according to visual estimation comparisons with a 1 kb ladder (Lane 10). Inconsistencies in the gel may have caused curvature in from endpoints, as is seen above. Pronounced primer dimerization is observed in Lane 1, 8, and 9, with minor amounts visibile in each other lane (2-7). Relevant fragment sizes on reference ladder are indicated in white to the right of each ladder in base pairs (bp).

After amplification of the PCR products from pGEMT-easy using RE template primers, RE digest and ligation were found to work successfully. Successful subcloning of this promoter region from pGEM-T easy into pGL4.23 was shown based on PCR detection of recombinants. Fourteen colonies were assayed using DreRyR3F1 and DreRyR3R5 and were found to be carrying the desired 1.8 kb fragment within the expression vector ligated after growing up in competent *E. coli* JM109 cells (Figure 7). These pGL4.23 vectors will be useful in future functional studies testing the transcriptional activity driven the promoter sequences within them.

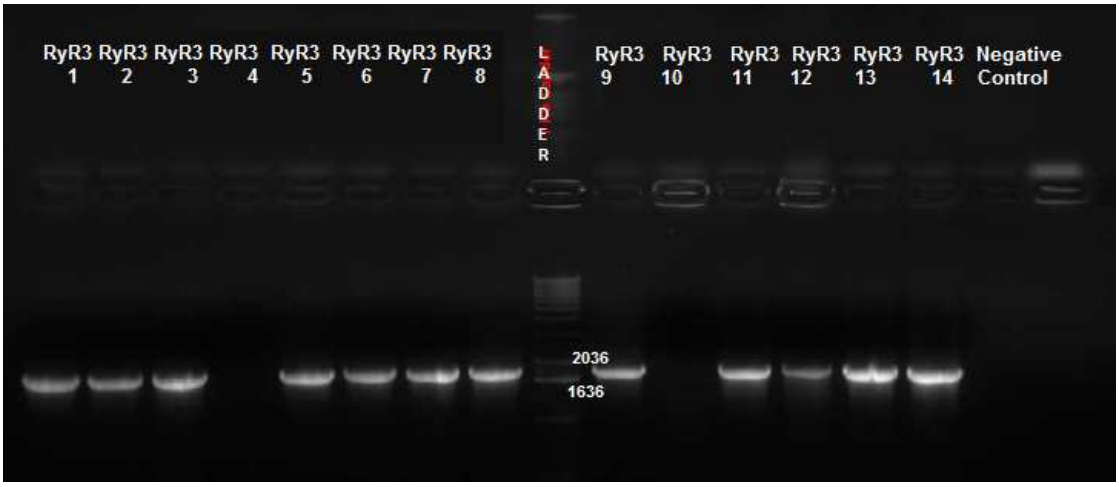


Figure 7: PCR detection of recombinants using primer pair DreRyR3F1 and DreRyR3R5. Products shown above were amplified from template *E. coli* JM109 colonies transformed with an expression vector, pGL4.23. Successful recombinants were detected in all Lanes except for Lanes 4 and 10. Lane 15 was a negative control, and a 1kb ladder is shown between Lanes 8 and 9.Detection of 2 kb fragments is only possible if expression vectors contain the upstream promoter sequence subcloned from pGEM-T easy. PCR product amplification was visualized on a 1% agarose gel stained with ethidium bromide. All products appear to be consistently within the size range of 1.8 kb, according to visual estimation comparisons with a 1 kb ladder Fragment sizes are shown at the central ladder in base pairs (bp).

All three sets of products obtained were purified and sent for sequencing, and assembled into a contig using the Lasergene Seqman Pro sequencing software (DNASTAR inc., Madison, Wi.) resulting in a contig with 6-fold coverage. The total length of the assembled contig was 1808 bp, compared to the same sequence in Genbank of 1809 bp (Appendix E). Comparison to the RyR3 Genbank sequence upstream of GeneID: 561350 identified 13 single nucleotide polymorphisms (SNPs; Table 2; Figure 8).

Table 2: A) Relative position of important SNPs within the upstream promoter region of RyR3, as indicated by ZV7 (Genbank; GeneID: 56150) compared to an assembled Contig, ZFRyR3. This alignment is represented graphically in Figure 8. Note the presence of 13 SNPs. **B)** Position of SNPs which are notably different between sequences. These may be deletions or ambiguous base pairs, as indicated.

A

Position	51	122	158	181	231	234	321	570	837	1052	1241	1290	1338
Contig	G	A	T	G	T	T	A	A	T	T	G	G	A
Genbank	A	C	G	A	A	A	G	T	A	A	A	A	C

B

Position	489	571	1007	1031
Contig	T	-	Y	-
Genbank	-	T	T	A

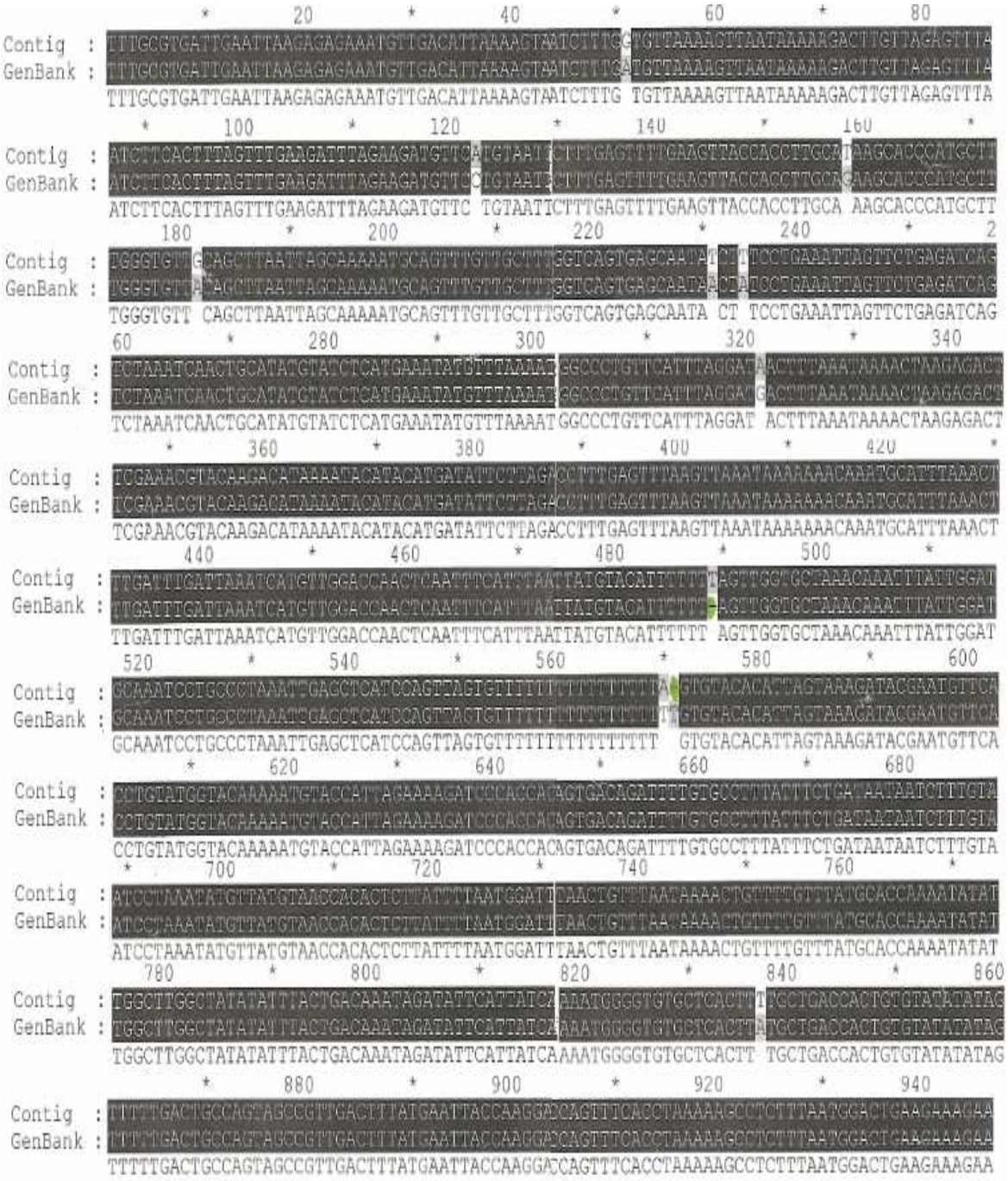


Figure 8: Nucleotide sequence alignment of a region upstream of RyR3 from ZV7 (GenBank; GeneID: 56150) and a Contig, ZFRYR3. This contig has been assembled from *D. rerio* clones using sequences amplified from *D. rerio* genomic DNA samples ZFG1, ZFG2, and ZFG4 as well as sequences taken from expression vectors subcloned from stored pGEM-T easy clones. Regions which are conserved are shaded black with a consensus sequence generated and appearing below each line. Nucleotide coordinates are shown above each line. A total of 1808 bp from ZFRYR3 and 1809 bp from GenBank are included. Note the occurrence of 13 SNP sites throughout the figure.

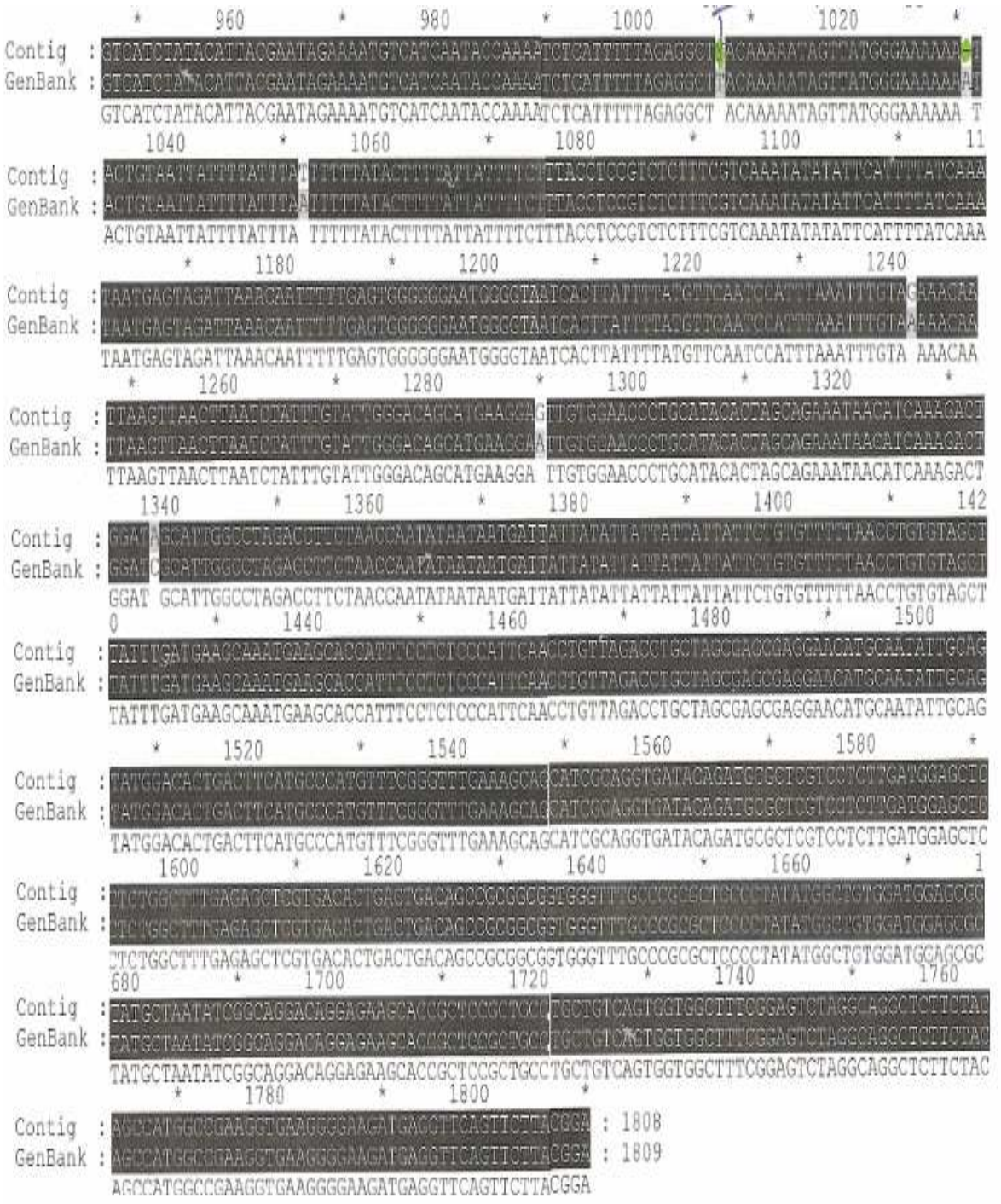
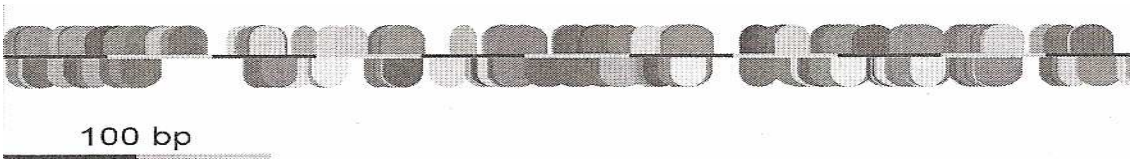


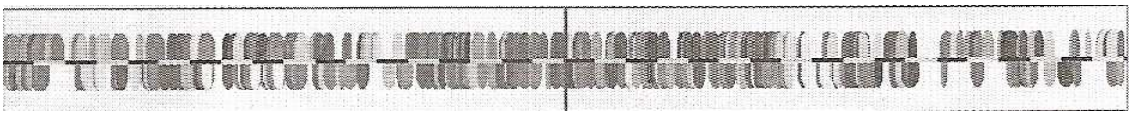
Figure 8: Nucleotide sequence alignment of a region upstream of RyR3 from ZV7 (GenBank; GeneID: 56150) and a Contig, ZFRYR3. (continued)

3.2 MatInspector Results

The RyR3 upstream promoter sequence was analyzed using the high throughput search engine, MatInspector (Genomatix). Using a precomputated vertebrate weight matrix, and performing a search for all TF site hits with a specificity to their core sequence of 0.75 (as recommended to minimize false positives while allowing for the revelation of binding sites with many permutations), an output image was generated (Figure 9). In total, 706 matches for TF binding sites were identified within this region. A list of these factors was generated, including their function, abbreviated identity, matrix utilized, and start position in comparison to a given bp anchor based on their predefined core sequence anchors.



A



B

Figure 9: General core promoter elements located to the RyR3 upstream promoter region using the MatInspector search engine. (A) 100 bp fragments are indicated by divided black and white lines throughout the small core segment shown (approximately 550 bp). Each half circle displayed represents a putative TF binding site. Half circles occuring below the sequence lines are overlapping TFBSs identified to occur at the same nucleotide coordinates. (B) View of the entire 1808 bp region based on ZFRYR3, showing a multitude of TFBDs.

Identified TFBSs included FOX factors, myogenin/myoD factor binding sites, and sites for the binding of factors from other TF families, including the CART family, NKXX

family, CLOX, and zinc finger motif proteins. The MatInspector database search results were screened again manually and factors with a matrix similarity of 0.85 or greater were selected from this list for further analysis. This filtering step resulted in 200 TFBS hits (not included). Given the large amount of site overlap, a second program search output was used to confirm TFBS type and location, using a cutoff of 0.90 similarity to a core sequence to minimize false positives and using a different matrix system. Thus, the results from MatInspector serve as comparisons to results from Match in determining the accuracy of TFBS hits.

3.3 Match Results

A search output for ZFRYR3 was next created using the program Match (Waleev *et al,* 2006). Two separate outputs were generated. In doing so, a muscle specific profile (muscle_specific.prf) was utilized to try and locate significant TFBSs of a muscle-specific nature (Table 3). A second output was also generated scanning with a vertebrate matrix group and a cut-off instruction to minimize false positives (i.e. a core match of 0.9 or greater and a matrix match of 0.85 or greater; Table 4). A variety of sites hits for TFs not identified by MatInspector were generated, including multiple five site hits for SP1, four site hits for USF, and two site hits for SRF. Match also identified several different sites as a possible TATA box, at a variety of locations. A frequency of sites per nucleotide was found using the muscle specific search profile of 0.013274, compared to a frequency of 0.003319 generated using the minimization of false positives cut-off. In general, this means that fewer hits were generated in the latter case, with a higher likelihood of each hit being accurate. After comparing obtained results from Match and MatInspector, a

list of TFBSs was generated to include TFBSs identified by both program searches, including some additional TFBSs in those families.

TABLE 3: TF site search results using Match (gene-regulation.com). A) Matrix library TFP60.lib was used, a vertebrate specific matrix library with a search parameter of muscle specific profile (muscle_specific.prf). Sequence results generated for ZFRYR3 are shown above, including a multitude of putative myoD sites. Matrix similarity refers to the similarity of each sequence position to the highest conserved nucleotide at that position in a given matrix. Core similarity refers to the similarity of a sequence to the 4 highest conserved positions in a matrix.

Matrix Identifier	Position (bp)	Core Match	Matrix Match
TATA	32	1.000	0.895
MYOD	263	0.973	0.833
TATA	292	1.000	0.900
TATA	358	1.000	0.886
SP1	520	0.819	0.787
E47	595	1.000	0.921
MYOD	597	1.000	0.877
USF	598	0.926	0.871
SP1	634	0.915	0.779
TATA	718	1.000	0.889
TATA	994	1.000	0.926
TATA	1051	1.000	0.947
MYOD	1114	0.778	0.811
SP1	1185	0.904	0.829
TATA	1210	1.000	0.886
USF	1405	0.918	0.865
E47	1547	1.000	0.946
MYOD	1549	1.000	0.918
USF	1550	0.926	0.880
USF	1603	1.000	0.943
SP1	1625	1.000	0.822
SP1	1628	0.915	0.786
SRF	1647	0.880	0.896
SRF	1650	1.000	0.929

Table 4: TF site search results using Match (gene-regulation.com). Core matches and matrix matches are used to minimize false positive matches. (B) Matrix library TFP60.lib was again used, with a cut-off employed to searching of ZFRYR3 in order to minimize false positive matches. Six sites were identified within this 1808 bp region, including a site for FOX domain binding, a site for OCT-1 binding, an HLF binding site, an HNF1 binding site, and an NKXX site. Matrix similarity refers to the similarity of each sequence position to the highest conserved nucleotide at that position in a given matrix. Core similarity refers to the similarity of a sequence to the 4 highest conserved positions in a matrix

Matrix Identifier	Position (bp)	Core Match	Matrix Match
FOXD3	408	1.000	0.951
OCT1	511	1.000	0.952
HLF1	699	1.000	0.945
VBP1	699	0.982	0.976
HNF1	800	0.920	0.873
NKX25	1243	1.000	1.000

Due to the vast number of site hits discovered by MatInspector, and the variety and high density of TFBSs identified, the choice was made to approach comparative analysis of RyR3 upstream promoter regions from various species using a larger 5' upstream region than 2 kb. As such, a 10,000 bp region was identified upstream of GeneID: 56150 from ZV7 and compared to the upstream region of RyR3 in fugu (*Takifugu rubripes*), tetraodon (*Tetraodon nigroviridis*), medaka (*Oryzias latipes*), and human (*Homo sapiens*). The identities of the RyR3 orthologues, located on the Ensembl genome browser, are as follows: *Takifugu rubripes* RyR3a ENSTRUG00000014204, RyR3b ENSTRUG00000018032; *Tetraodon nigroviridis* RyR3a ENSTNIG00000012214, RyR3b ENSTNIG00000012216, RyR3c ENSTNIG00000017445; *Oryzias latipes* RyR3a ENSORLG00000016893, RyR3b ENSORLG00000017871; and *Homo sapiens* RyR3 ENSG00000198838. In cases of alternative splicing, sequences were taken starting from the start site of transcription furthest upstream.

3.4 mVISTA Results

Initial analysis using mVISTA showed very little conservation between the *H. sapiens* and *D. rerio* upstream promoter regions; this was confirmed using ClustalX 2.0 (not included; Figure 10). Sequences were compared using a Shuffle-LAGAN system, which relies on global alignment of finished sequences. Rearrangements and inversions may be detected by this program, maximizing alignment possibilities. mVISTA produces a global end-end map in this case, while enabling one to look at all pair-wise combinations. Despite the relatively low degree of sequence conservation observed, pairwise sequence comparisons were saved and submitted for further analysis using the

rVISTA tool. This was saved to search for TFBS hits within conserved regions and aligned regions between promoters.

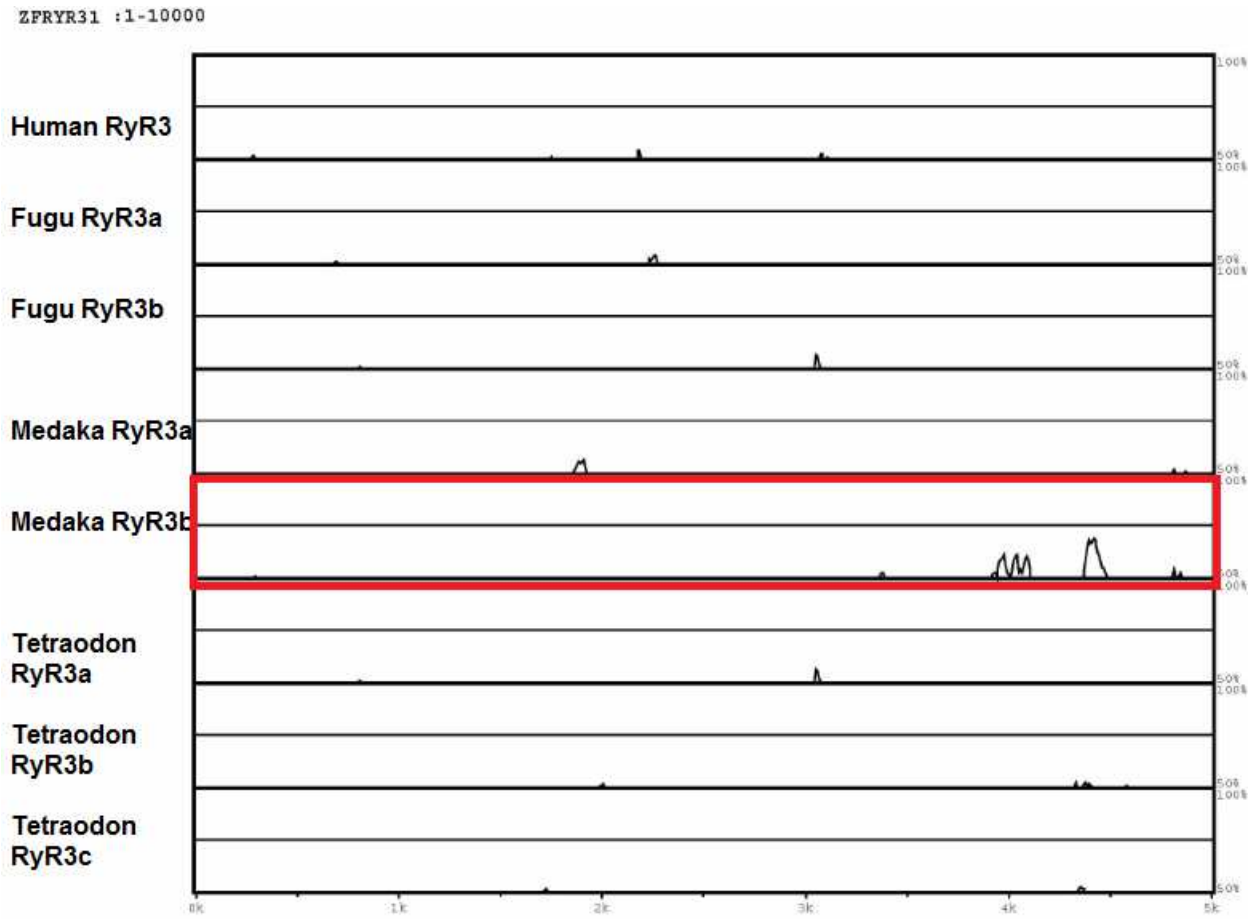


Figure 10: VISTA browser output for sequence conservation between a 10,000 bp promoter sequence of RyR3 in *D. rerio* and 8 other RyR3 promoter sequences. Sequence comparisons were performed using the mVISTA program with a default 100 bp sliding window. On the x-axis, sequence coordinates are shown in the positive direction; thus, a site indicated at 1 k above is 9 k from the transcription start site. Regions shaded in orange correspond to sequence fragments with >70% sequence identity to the 10,000 bp upstream RyR3 *D. rerio* sequence-no such regions are shown. Very little conservation is shown above, with the highest degree of sequence similarity observed with MedakaRyR3b (highlighted in red). No points of 70% conservation were observed within 10 kb of the transcription start site between *D. rerio* and any of 8 RyR3 genes from *T. rubripes, T. nigroviridis, O. latipes,* or *H. sapiens.*

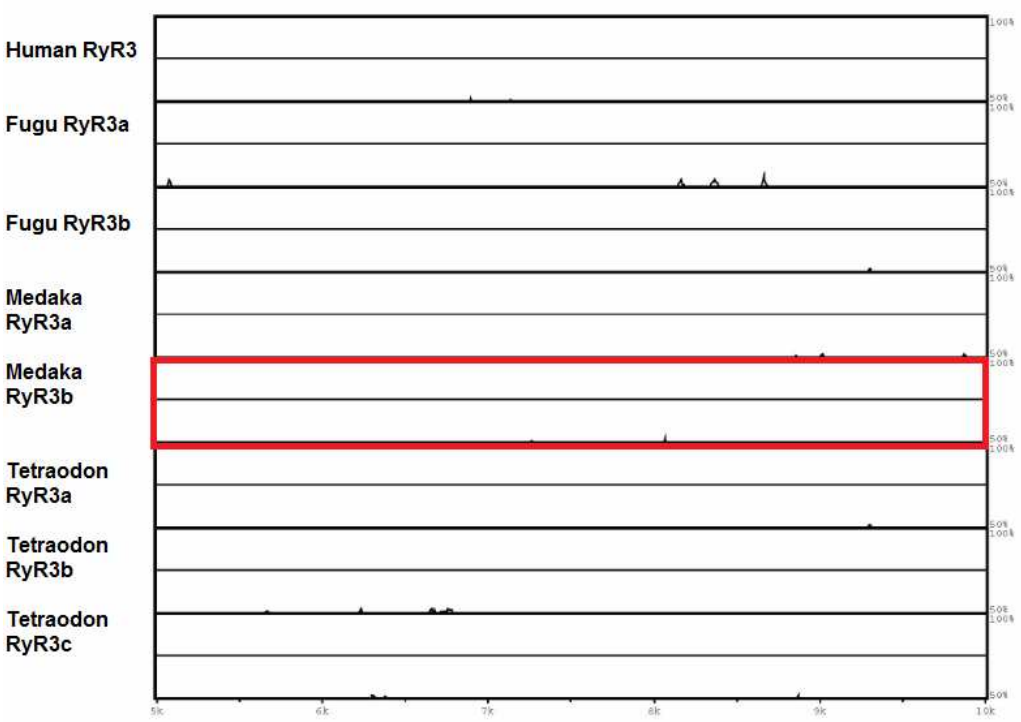


Figure 10: VISTA browser output for sequence conservation between a 10,000 bp promoter sequence of RyR3 in *D. rerio* and 8 other RyR3 promoter sequences. (continued)

3.5 rVISTA Results

Pairwise sequence comparisons from mVISTA were submitted for rVISTA analysis in order to search for TFBS hits located at any given point in either of the pair of promoter sequences between zebrafish RyR3 and each other RyR3. Furthermore, a second round of analysis was performed to find conserved and aligned TFBS regions between each pair. A group of TFs were chosen for searching of 455 available TFs based upon two initial criteria. 1) The TFBS must have been located by MatInspector and have had matrix identity of 0.85 or greater in the initial 2 kb analyzed. 2) The TFBS must have been located in one of two rounds of MATCH database searching as well, using the Transfac database output to increase reliability. In addition, some TFBSs were chosen

for searching due to their similarity to a TF that met the original two criteria (i.e. HNF3α). The following list of TFBSs was used for searching based on these criteria: FOX, MEF-2, MUSCLE, MYOD, MYOGENIN, STAT, CART-1, CLOX, EVI-1, GATA, HANDIE47, HNF-1, HNF-3, HNF-3α, HNF-4, HNF4-α, HNF-6, MYC, MYCMAX, MYOGNF-1, NKX6-1, NKX6-2, OCT, and TATA (Abbreviations and Nomenclature). MUSCLE is a core sequence which takes the bHLH domains of many muscle-specific TFBSs and decreases stringency for flanking sequences in order to produce a greater number of hits. In all pairwise comparisons, hundreds of hits were found using a 24 bp sliding window. This result was obtained after setting clustering to 0, an option which results from instruction of searching for 1 site over 100 bp. One such rVISTA output is shown below, although a majority of the image has been omitted due to its large size (Figure 11). Clustering was observed of EVI1 and HNF sites, while relatively few sites for muscle-specific transcription factors were observed. Other TFBSs were found to have hits in a relatively distributed fashion, including OCT, TATA, FOX, HANDIE47, and GATA. CART-1 and CLOX hits were also abundant. A search for TFs which are both conserved and aligned was then performed using *D. rerio* promoter as a baseline in each pairwise comparison, as previously indicated, resulting in the eight different outputs. VISTA offers the ability to discern between TFs that are aligned, and those that are aligned and within conserved regions. With the exception of Medaka RyR3b, no aligned or conserved TFBSs were found within the first 2 kb of any promoter region comparison. Additionally, Medaka RyR3b contained sites which were aligned in the first 2 kb of the promoter region, but not conserved. Within other regions of conservation as determined by mVISTA (Figure

11), TF sites were discovered that were aligned, but those were not observed until at least 2500 bp into the sequence (i.e. 7500 bp coordinates; Figure 12; Table 5). Comparisons to RyR3a, b, and c in *T. nigroviridis,* RyR3a in *O. latipes,* and RyR3a, and 3b in *T. rubripes* found no TF sites which were both aligned and within a conserved region. These results were reconfirmed in comparisons with *H. sapiens* RyR3 promoter (Figure 14). A consistently low abundance of muscle-specific transcription factor hits was observed in all cases, and an abundance of HNF hits was found. Most significantly, HNF4 was found in high abundance in clusters in many conserved regions. These sites may be of functional importance, and should be investigated in future expression studies.

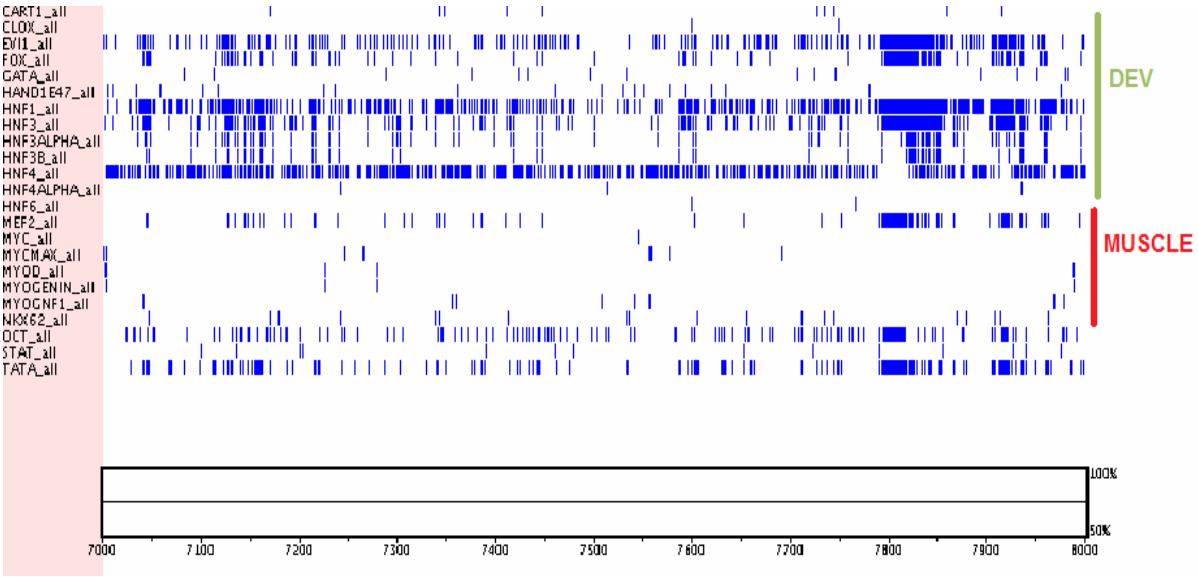


Figure 11: rVISTA output for TFBS hits between a region 2,000-3000 bp upstream of RyR3 in *D. rerio* and *H. sapiens*. Site seeking was performed using the rVISTA program with a default 24 bp sliding window and a threshold of 0.85 match to core sequence. On the x-axis, sequence coordinates are shown in the positive direction; thus, a site indicated at 7000 bp above is 3000 bp from the transcription start site. On the left, the names of each TF are written, with the designation "all" referring to the search for all hits of a given TF site in either sequence, regardless of conservation or alignment. Blue lines parallel to TF names indicate such a hit. Note the relatively low abundance of muscle-specific TF hits (red bolded line), while a great abundance for HNF sites is observed (green bolded line; dev=development). Clustering of TFBSs is observed in this small sampling of a 10,000 bp region for HNF and EVI1 sites.

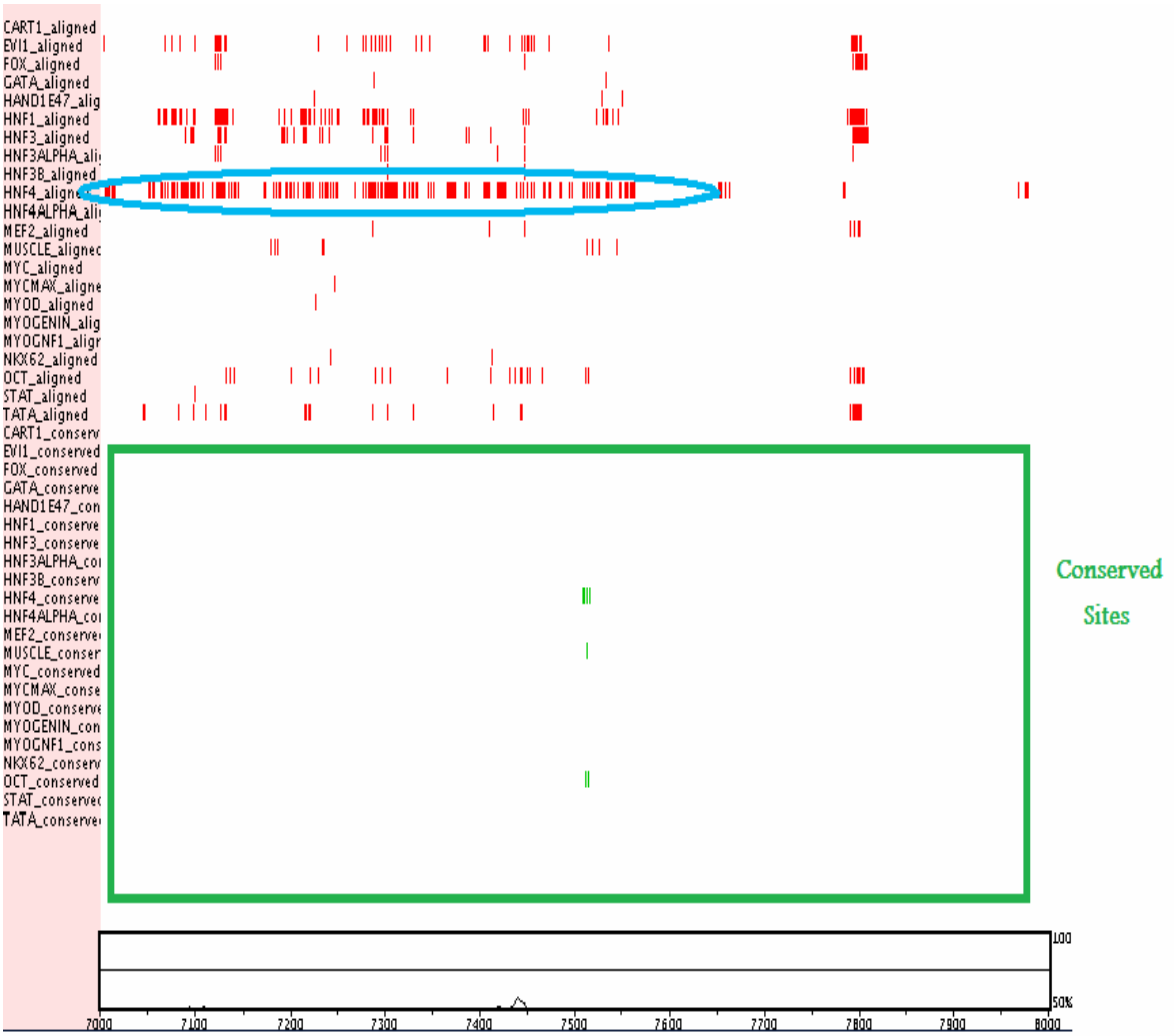


Figure 12: rVISTA output for conserved and aligned TFBS hits between a region 2,000-3000 bp upstream of RyR3 in *D. rerio* and *O. latipes* RyR3b. Site seeking was performed using the rVISTA program with a default 24 bp sliding window. On the x-axis, sequence coordinates are shown in the positive direction; thus, a site indicated at 7500 bp above is 2500 bp from the transcription start site. On the left, the names of each TF are written, with the designation "conserved" referring to the search for hits of a given TF site in the conserved region of the pair of sequences, indicated by green lines parallel to TF names. The designation "aligned" refers to results from a search for TF hits where TFs fall in alignment for both sequences, indicated by Red lines. Note the relatively low abundance of muscle-specific TF hits, while a great abundance for HNF sites is observed (teal circle above). Clustering of TFBSs is observed in this small sampling of a 10,000 bp region for HNF. Four HNF-4, one MUSCLE, and two OCT sites are shown to be conserved and aligned (indicated in green).

Table 5: Transcription factor binding site hits for conserved and aligned TFs within a pairwise comparison of the RyR3 upstream promoter region of MedakaRyR3b (*O. latipes,* ENSORLG00000017871, (using a 10,000 bp upstream region of *D. rerio* RyR3 as a baseline.)*

Transcription Factor Name	TF Type	Approximate Coordinates	Number of Sites
OCT	General	7500 bp	2
		5050 bp	2
		4000 bp	1
		4050 bp	1
		4400 bp	1
		2300 bp	1
MUSCLE	Muscle	7500 bp	1
		4100 bp	1
HNF-4	Developmental	7500 bp	3
		5050 bp	3
		4000 bp	2
		4050 bp	C
		4100 bp	C
		4350 bp	C
		4450 bp	C
		2300 bp	C
TATA	General	5050 bp	2
		4000 bp	1
		2300 bp	1
MEF-2	Muscle	5050 bp	1
HNF-3	Developmental	5050 bp	1
		4100 bp	C
		4300 bp	C
HNF-3α	Developmental	5050 bp	1
		4300 bp	1
		4100 bp	1
EVI-1	Developmental	5050 bp	2
		4050 bp	1
		2300 bp	1
Myogenin	Muscle	4100 bp	1
		4400 bp	1
FOX	Developmental	4100 bp	1
MyoD	Muscle	4100 bp	1
		4400 bp	1
MyCmax	Muscle	4050 bp	2
		4100 bp	1
HNF-4α	Developmental	4400 bp	1
		4100 bp	1
HNF-1	Developmental	4400 bp	C
		4100 bp	1
		4050 bp	2
HAND1E47	Developmental	4450 bp	1
GATA	Developmental	4400 bp	1
		4450 bp	1

* Coordinates of hits are approximations only. The letter C is used to represent a large cluster of sites for the same TF in the same location in a conserved region.

4. Discussion

4.1 Bioinformatics Data

Analysis and characterization of the regulation of gene expression poses several challenges. In essence, certain TFBS hits which were encountered throughout searching are puzzling. For instance, TATA boxes are usually found upstream of the translation start site of a given gene, but not in multiples, a feature present in the RyR3 promoter in *D. rerio*. In certain species including humans, different RyR3 transcripts are produced through alternative splicing; the location of these TATA box hits, however, does not closely correspond with the locations of these alternative splice sites (ENSG00000198838). Additionally, for example, ectopic viral integration protein 1 (EVI-1) is a proto-oncogene involved in maintenance of tumor suppression and promotion of differentiation from bone marrow progenitor cells to granulocytes (Weiser 2007). Without molecular biological experimentation to explore the role of RyR3 in either of these processes, it is difficult to determine the significance of such a find. It is therefore difficult to understand the validity of this bioinformatics data, although some key pieces of data may be significant. The overall message, then, is that it is important to consider that potential binding site identification may be misleading.

As a prime example, the discovery of clusters of HNF sites within a 10,000 bp upstream region of RyR3 in *D. rerio* with a noticeably lower abundance of muscle-specific sites is intriguing. The hepatocyte nuclear factors (HNF) are members of the winged helix family of TFs, involved in a variety of functions (Jacob *et al,* 1997). Previous studies have shown examples of HNFs used as a transcriptional activator of

transthyretin, a serum carrier protein for Vitamin A and thyroid hormone (Li *et al,* 2000). Additional roles for this group include hepatocyte differentiation, and initial differentiation from endoderm cells (Divine *et al,* 2003). While these roles do not seem related to the classical function of the RyR, this may indicate that RyR3 plays some role in the development of other non-contractile tissues such as the liver. One other reason for such a finding may be related to the endogenous expression of RyR3 in brain tissue within the zebrafish. Certain HNF factors such as HNF-6 are expressed in developing zebrafish brain, including expression in forebrain and midbrain at 12 hpf; functional studies will be required to explore these possibilities (Matthews *et al,* 2004).

While it was initially surprising to find such an abundance of *cis*-regulatory elements, closer analysis of which TFs cluster and occur most often reveals potential functional significance. Previous work on experimentally determined muscle regulatory sequences indicates that muscle expression requires multiple elements in close proximity (Wright *et al,* 1991). This is in line with studies of the HNF factors and the role they play in the regulation of the FabpI gene which is expressed in intestinal and hepatic cells (Divine *et al,* 2003). In this circumstance, *Fabpl* transgene was directly activated through enhancer sites via synergistic reactions of HNF-1α HNF-1β, GATA-4, FoxA2, and HNF-4α (Sladek *et al,* 1990). Strikingly, HNF-1α activated the *Fabpl* transgene by as much as 60-fold in the presence of the other five endodermal factors than in their absence, accounting for up to one-half of the total transgene activation by the group of six factors (Divine *et al,* 2003). This suggests that other TF sites in conserved regions occuring near HNF sites, including the FOX and GATA sites discovered in the regulatory

region of RyR3, may play a role in synergistically activating RyR3 transcription (Li *et al,* 2000, Jacob *et al,* 1997). Additionally, sites of related transcription factors concentrated in a given area have been observed in previous studies based on globin genes, muscle-specific genes, and cell-cycle dependant genes due to their concerted role in the regulation of those genes (Dermitzakis and Clark 2002). Thus, these clustered sites may play a key role in RyR3 regulation in fish.

OCT-1 is a type of TF that is involved in regulating the expression of tissue-specific and housekeeping genes (Lee *et al,* 2001). The interaction of OCT-1 with other factors is complex. Lee and colleagues showed no physical interaction between OCT-1 and BoB1 to stimulate immunoglobulin gene development, even though the presence of the two factors clearly activates transcription synergistically (2001). These authors posit that factors are co-recruited and are used in groups to form a vast regulatory assembly. Thus the presence of this factor in conserved regions of the RyR3 promoter region may be related to a role in the development of a variety of tissues.

In terms of muscle-specific TFBSs, the discovery of only nine conserved and aligned elements between RyR3B in *O. latipes* and *D. rerio* was unexpected. Transfection into undifferentiated ZF4 cell lines will eventually help in the resolution of this issue, but further complication arises from the overlap of muscle-specific sites such as MEF2 with developmental sites such as FOX. Deletion of these regions using convenient RE sites may lead to a raised or lowered transcriptional activity, allowing for better resolution of the potential implications for spatiotemporal expression of RyR3.

4.2 Regulatory DNA Region Characterization

Most of the difficulty in characterizing regulatory DNA regions is multifactorial. First and foremost, TFBSs are difficult to characterize properly due to their ability to combine in functional combinations (Vavouri and Elgar 2005). Scientists have struggled to create a list of properties by which to characterize these regions. In coding sequences, open reading frames (ORF) and non-uniform codon usage exist to help in characterization (Abnizova and Gilks 2006). Since there are no well characterized properties of regulatory DNA analogous to these, it makes it nearly impossible to define a consensus and location of regulatory elements (Abnizova and Gilks 2006). Secondly, TFs have low specificity for their binding site motifs, making it difficult to accurately detect sites (Vavouri and Elgar 2005). This is the main reason why setting filters for TFs is so complex; weight matrices are statistical in nature, and do not account for differing binding site preferences based on flanking sequences (Waleev *et al,* 2006). While a smaller sliding window may reveal a different number of TFBS hits in programs such as rVISTA, I believe that the use of different bioinformatics programs such as EBioX (Ebioinformatics) may assist in the resolution of some of these issues, because search outputs from a different program with a different set of parameters are useful as confirmation tools. Using a different group of matrices in searching may at least reinforce the significance of TFBS hit findings, if they are to reoccur.

Differences in regulatory regions in different genes also cause a major issue in drawing inferences, as promoters in general comprise a collection of diverse TFBSs with an enormously varying composition and organization (Abnizova and Gilks 2006). Further

to this, TFBSs are dispersed unevenly; in summation, there are no regular structural elements of regulatory regions (Wasserman *et al,* 2000). Adding to this issue is the major cost and time resource consumed by experimental verification of regulatory region elements. Since it is only possible to draw inferences from *In silico* results, molecular biological experimentation is absolutely necessary for confirmation, including the use of DNAse I footprinting, chromatin Immunoprecipitation (ChIP) and other techniques that allow for visualization of different protein-DNA complexes (Blanchette and Tompa, 2002; Drivenes *et al,* 2000).

4.3 RyR3 Promoter Comparisons

Regulatory element evolution does not have a clearly established set of rules, because we do not yet have a complete handle on the evolution of transcription regulation (Abnizova and Gilks 2006). This leads to difficulty when comparing promoter sequences between species, as was performed in mVISTA. Chang and colleagues encountered similar difficulties in summarizing their findings, working with a pufferfish model and GFP expression assays (2006). While I found very few sequence-specific relationships in comparing a 10,000 bp region upstream of RyR3 in *D. rerio* to the same region in several other mammalian and non-mammalian vertebrate species, the finding of great similarity between the promoter of RyR3b in *O. latipes* and RyR3 in *D. rerio* is interesting. The last common ancestor of *O. latipes* and *D. rerio* lived ~140 million years ago, and this ancestor is believed to be the common ancestor of almost all euteleosts (Naruse *et al,* 2004). In experiments designed to study ancestral gene function, the spectrum and phenotypes of induced mutations in *O. latipes* is often different from

those found in *D. rerio* (Ishikawa 2000; Loosli et al. 2000). Further to this, while minimal sequence conservation was observed between *H. sapiens* and *D. rerio* RyR3 promoter, studies performed using the gene for receptor tyrosine kinase, *RET*, have shown that the upstream noncoding sequence in teleosts such as *D. rerio* is capable of driving transcription of *RET* in humans (and vice versa; Fisher *et al,* 2006). There is a clear need to test whether RyR3 promoter from *D. rerio* can drive transcription of orthologous RyR3s. If this is the case, it may be expected that RyR3 promoter from *O. latipes* or *H. sapiens* could drive *D. rerio* RyR3 transcription. This work is essential in understanding the relationship between such promoter regions.

4.4 Future Work

In order to clarify the results of *In silico* analysis, future work will include the transfection of the pGL4.23 expression vectors into *D. rerio* cell lines of different types. These cell types will vary in characteristics that will allow for the exploration of promoter activity in temporally and in spatially distinct ways. Other functional assays including ChIP will help in visualizing which TFs truly bind to the promoter, allowing for better resolution of the function of RyR3 in *D. rerio*.

5. Conclusions

1. Results of comparison of a 10,000 bp region upstream of the RyR3 gene in zebrafish, *Danio rerio,* reveal a multitude of potential binding sites for transcription factors (TF) related to development, including clusters of HNF family sites, OCT sites, and EVI-1 sites.
2. A lower abundance of muscle-specific transcription factor binding sites (TFBS), including those for myogenin, myoD, and MEF-2, was observed compared to binding sites for developmental transcription factors.
3. The highest degree of sequence conservation was observed between the 10,000 bp promoter region of Medaka (*Oryzias latipes*) RyR3b, and that of *D. rerio.* No conservation of sequence of >70% was observed between this sequence in *D. rerio* and the same sequence in any RyR3 promoter for *Tetraodon nigroviridis, Takifugu rubripes, Oryzias latipes,* and *Homo sapiens.*
4. Further molecular biological studies, including ChIP and transfection assays, will be necessary to confirm the findings of functional promoter elements.

6. References

Abnizova, I., and W.R. Gilks. 2006. Studying statistical properties of regulatory DNA sequences, and their use in predicting regulatory regions in the eukaryotic genomes. Briefings in Bioinformatics 7 (**1):** 48-54.

Ashworth, R., Zimprich, F. and S.R. Bolsover. 2001. Buffering intracellular calcium disrupts motoneuron development in intact zebrafish embryos. Developmental Brain Research 129: 169-179.

Ashworth, R. and S.R. Bolsover. 2002. Spontaneous activity-independent intracellular calcium signals in the developing spinal cord of the zebrafish embryo. Developmental Brain Research 139: 131-137.

Ashworth, R. 2004. Approaches to measuring calcium in zebrafish: focus on neuronal development. Cell Calcium 35: 393-402.

Ashworth, R., Tunwell, R., Koh, K.R., and S. Patel. 2007. Role of IP3 receptors during early zebrafish development. Journal of Biological Chemistry 282 **(19):** 13984-93.

Bers, D.M. 2002. Cardiac excitation-contraction coupling. Nature 415 (**6868):** 198-205.

Blanchette, M., and M. Tompa. 2002. Discovery of regulatory elements by a computational method for phylogenetic footprinting. Genome Research 12 **(5):** 739-748.

Brennan, C. and E.N. Olson. 1990 Myogenin resides in the nucleus and acquires high affinity for a conserved enhancer element on heterodimerization. Genes and Development 4:582-595.

Brennan, C., Mangoli, M. Dyer, C.E. and R. Ashworth. 2005. Acetylcholine and calcium signaling regulates muscle fibre formation in the zebrafish embryo, *Journal of Cell Science* 118**:** 5181–5190.

Brunet, A., Sweeney, L.B., Sturgill, J.F., Chua, K.F., Greer, P.L., Lin, Y., Tran, H., Ross, S.E., Mostoslavsky, R., Cohen, H.Y., Hu, L.S., Cheng, H.L., Jedrychowski. M.P., Gygi, S.P., Sinclair, D.A., and M.E Greenberg. 2004. Stress-dependent regulation of FOXO transcription factors by the SIRT11 deacetylase. Science 303(5666): 2011-2015.

Chang, M., Chou, C., Hsieh, Y., Lu, I., Devi, M.K.N., Chang, J., Kuo, T., and C. Huang. 2004. Identification of 5'-upstream region of pufferfish ribosomal protein L29 gene as a strong constitutive promoter to drive GFP expression in zebrafish. Biochemical and Biophysical Research Communications 314: 249-258.

Davis, R.L., Cheng, P.F., Lassar, A.B., and H. Weintraub. 1990. The MyoD DNA binding domain contains a recognition site for muscle-specific gene activation. Cell 60: 733-746.

Dermitzakis, E.T., and A.G. Clark. 2002. Evolution of transcription factor binding sites in mammalian gene regulatory regions: conservation and turnover. Molecular biology and Evolution 18: 114-1121.

Darbandi, S. and J.P. 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). Comparative Biochemistry and Physiology: Part B, Biochemistry and Molecular Biology 154 (**4):** 443-448.

Divine, J.K., McCaul, S.P., and T.C. Simon. 2003. HNF-1α and endodermal transcription factors cooperatively activate Fabp1:MODY3 mutations abrogate cooperativity. American Journal of Physiology Gastrointestinal Liver Physiology 285: G62-72.

Drivenes, O., Seo, H.C., and F. Anders. 2000. Characterization of the promoter region of the zebrafish *six7* gene. Biochimica et Biophysica Act 1491: 240-247.

Dubchak, I., and D.V. Ryabov. 2006. VISTA family of computational tools for comparative analysis of DNA sequences and whole genomes. Methods in Molecular Biology 338: 69-89.

Edmondson, D.G., Cheng, T., Cserjesi, P., Chakraborty, T., and E.N. Olson. 1992. Analysis of myogenin promoter reveals an indirect pathway for positive autoregulation mediated by the muscle-specific enhancer factor MEF-2. Molecular and Cellular Biology 12 **(9):** 3665-3677.

Fill, M. and J.A. Copello. 2002. Ryanodine receptor calcium release channels. Physiology Reviews 82 **(4):** 893-922.

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

Franck, J.P. C., Morrissette, J., Keen, J.E., Londraville, R.L., Beamsly, M., and B.A. Block. 1998. Cloning and characterization of fiber type-specific ryanodine receptor isoforms in skeletal muscles of fish. American Journal of Physiology 275: C401-C415.

Giannini, G.A., Conti, S., Mammarella, M., Scrobogna and V. Sorrentino. 1995. The ryanodine receptor/calcium channel genes are widely and differentially expressed in murine brain and peripheral tissues. Journal of Cell Biology 128 **(5):** 893-904.

He, S., Salas-Vidal, E., Rueb, S., Krens, S.F.G., Meijer, A.H., Snaar-jagalska, B. E., and H.P. Spaink. 2006. Genetic and Transcriptome characterization of model zebrafish cell lines. Zebrafish 3 **(4):** 441-452.

Ishikawa, Y. 2000. Medakafish as a model system for vertebrate developmental genetics. Bioessays 22: 487-495.

Jacob, A., Budhiraja, S., and R.R. Reichel. 1997. Differential induction of HNF-3 transcription factors during neuronal differentiation. Experimental Cell Research 234 **(2):** 277-284.

Karp, G. 2008. Cell and Molecular Cell Biology: Concepts and Experiments. 5th ed. New Jersey: John Wiley & Sons.

Koulen, P., Janowitz, T., Johenning, F.W., and B.E. Ehrlich. 2001. Characterization of the calcium-release channel/ryanodine receptor from zebrafish skeletal muscle. Journal of Membrane Biology 183 **(3)**: 155-163.

Lee, L., Stollar, E., Chang, J., Grossmann, J.G., O'Brien, R., Ladbury, J., Carpenter, B., Roberts, S., and B. Luisi. 2001. Expression of the Oct-1 transcription factor and \ characterization of its interactions with the Bob1 coactivator. Biochemistry 40 (**22):** 6580-6588.

Lehmann, O.J., Sowden, J.C., Carlsson, P., Jordan, T., and S.S. Bhattacharya. 2003. Fox's in development and disease. Trends in Genetics 19 **(6):** 339–344.

Li, J., Ning, G., and S.A. Duncan. 2000. Mammalian hepatocyte differentiation requires the transcription factor HNF4-α. Genes and Development 14: 464-474.

Loosli, F., Koster, R.W., Carl, M., Kuhnlein, R., Henrich, T., Mucke, M., Krone, A., and J. Wittbrodt..2000. A genetic screen for mutations affecting embryonic development in medaka fish (Oryzias latipes). Mechanisms of Development. 97: 133-139.

Mattei, M.G., Giannini, G., Moscatelli, F., and V. Sorrentino. 1994. Short communication: chromosomal localization of murine ryanodine receptor genes RyR1, RyR2, and RyR3 by *in Situ* hybridization. Genomics 22: 202-204.

Matthews, R.P., Lorent, K., Russo, P., and M. Pack. 2004. The zebrafish onecut gene hnf-6 functions in an evolutionarily conserved genetic pathway that regulates vertebrate biliary development. Developmental Biology 274 **(2):** 245-259.

Nabhani, T., Zhu, X., Simeoni, I., Sorrentino, V., Valdivia, H.H., and J. Garcia. 2002. Imperatoxin A enhances Ca^{2+} release in developing skeletal muscle containing ryanodine receptor type 3. Biophysical Journal 82: 1319-1328.

Naruse, K., Tanaka, M., Mita, K., Shima, A., Postlethwait, J., and H. Mitani. 2004. A Medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping. Genome Research 14: 820-828.

Nishida, 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. European Journal of Biochemistry 240: 408-415.

O'Brien, J., Meissner, G., and B.A. Block. 1993. The fastest contracting muscles of nonmammalian vertebrates express only one isoform of the ryanodine receptor. Biophysical Journal 65: 2418-2427.

O'Brien, J., Valdivia, H.H., and B.A. Block. 1995. Physiological differences between the α and β ryanodine receptors of fish skeletal muscle. Biophysical Journal 68: 471-482.

Paolini, C., Protasi, F., and C. Franzini-Armstrong. 2004. Evidence for conformational coupling between two calcium channels. Proceedings of the National Academy of Science of the United States of America 101 **(34):** 12748-12752.

Perez, C.F., Lopez, J.R., and P.D. Allen. 2005. Expression levels of RyR1 and RyR3 control resting free Ca^{2+} in skeletal muscle. AJP Cell Physiology 288: C640-C649.

Protasi, F., Paolini, C., Nakai, J., Beam, K.J., Franzini-Armstrong, C., and P.D. Allen. 2002. Multiple regions of RYR1 mediate functional and structural interactions with a1s DHPR in skeletal muscle. Journal of Biophysiology 83: 3230-3244.

Sawada, K., Hosoi, E., Bando, M., Sakata-Haga, H., Lee, N.S., Jeong, Y.G., and Y. Fukui. 2008. Different alterations in expressions of ryanodine receptor subtype in cerebellar cortical neurons of an ataxic mutant, rolling mouse nagoya. Neuroscience 152: 609-617.

Sladek, F.M., Zhong, W.M., Lai, E., and J.E. Darnell. 1990. Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. Genes and Development 4 (**12B):** 2353:2365.

Stem, D.L. 2000. Perspective: evolutionary developmental biology and the problem of variation. Evolution, 54:1079-1091.

Stuart, G., Vielkind, J., McMurray, J. and M. Westerfield. 1990. Stable lines of transgenic zebrafish exhibit reproducible patterns of transgene expression. Development 109: 577-584.

Stutzmann, G.E., Smith, I., Caccamo, A., Oddo, S., LaFerla, F.M., and I. Parker. 2006. Enhanced ryanodine receptor recruitment contributes to calcium disruptions in young, adult, and aged Alzheimer's disease mice. The Journal of Neuroscience 26 **(19):** 5180-5189.

Thibault, O., Gant, J.C., and P.W. Landfield. 2007. Expansion of the calcium hypothesis of brain aging and Alzheimer's disease: minding the store. Aging Cell 6: 307-317.

Vavouri, T., and G. Elgar. 2005. Prediction of *cis*-regulatory elements using binding site matrices – the success, the failures and the reasons for both. Current opinion on Genetic Development 15: 395-402.

Waleev, T., Shtokalo, D., Konovalova, T., Voss, N., Cheremushkin, E., Stegmaier, P., Kel-Margoulis, O., Wingender, E., and A. Kel. 2006. Composite module analyst: identification of transcription factor binding site combinations using genetic algorithm. Nucleic Acid Research 34**(web issue):**W541-W545.

Ward, C.W., Protasi, F., Castillo, D., Wang, Y., Chen, S.R.W., Pessah, I.N., Allen, P.D., and M.F. Schneider. 2001. Type 1 and type 3 ryanodine receptors generate different Ca^{2+} release event activity in both intact and permeabilized myotubes. Biophysical Journal 81: 3216-3230.

Wasserman, W., Palumbo, M., Thompson, W., Fickett, J.W., and C.E. Lawrence. 2000. Human-mouse genome comparisons to locate regulatory sites. Natural Genetics 26 **(2)**: 225-228.

Wieser, R. 2007. The oncogene and developmental regulator EVI1: expression, biochemical properties, and biological functions. Gene 396 **(2):** 346-357.

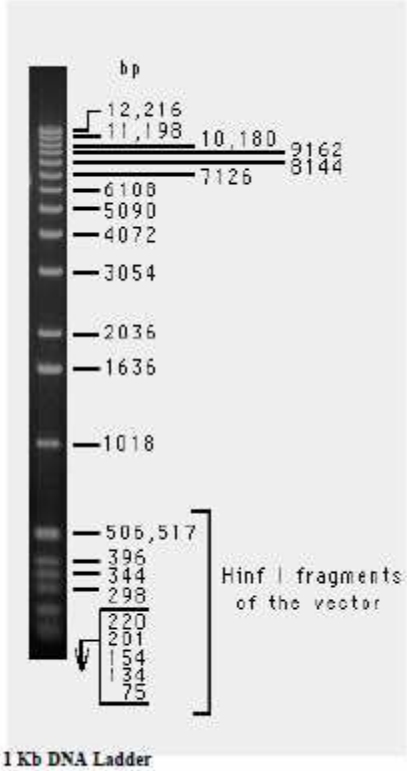
Wingender, E. 2000. The TRANSFAC system on gene regulation. Trends in Glycoscience and Glycotechnology 12: 255-264. <http://www.gene-regulation.com/pub/databases.html.>

Wright, W.E., Binder, M., and W. Funk. 1991. Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. Molecular Cell Biology 11:4104-4110.

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. Journal of Biological Chemistry 265 (**4):** 244-2256.

7. Appendices

Appendix A: 1Kb DNA Ladder (Invitrogen Life Technologies)



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

Structure of Fragment (1):

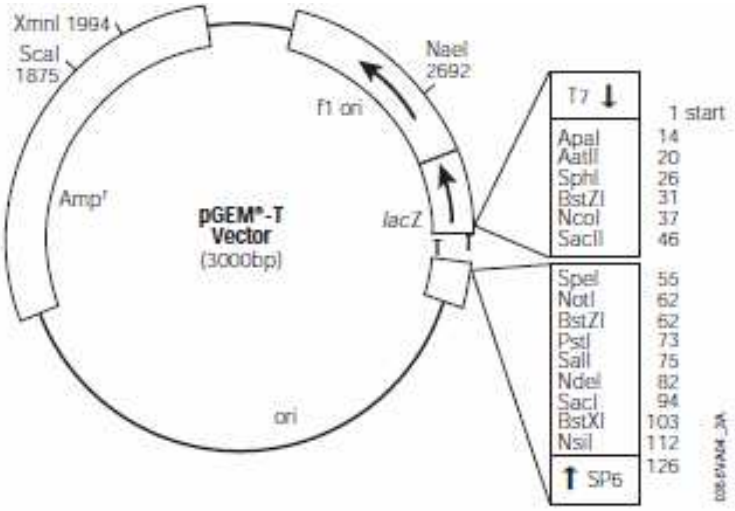
5′-TCGGG———C
C———GAGCC-5′
1014 bp

Notes:
During 1.2% agarose gel electrophoresis with Tris-acetate (pH 7.6) as the running buffer, bromophenol blue migrates together with the 506/517 bp doublet band.

The 1636 bp band and all bands less than 1000 bp are generated from pBR322.

If the ionic strength of the sample is too low, blurring of the bands can occur.

Appendix B: pGEM-T Easy Vector System Map (Promega Biotech)

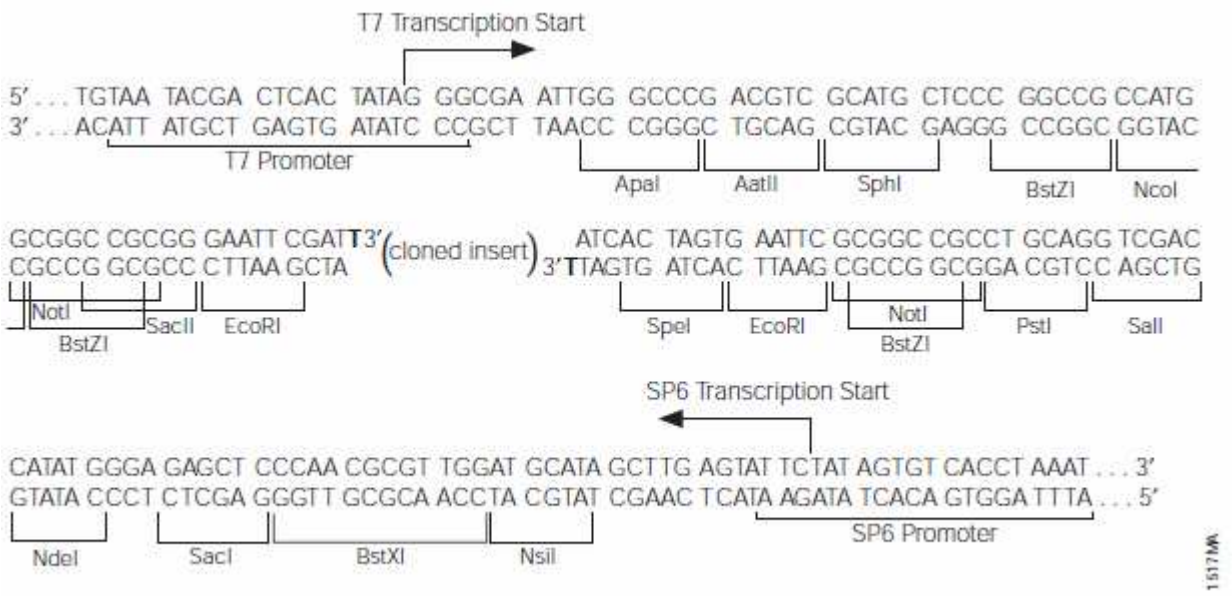


pGEM®-T Vector sequence reference points:

T7 RNA polymerase transcription initiation site	1
multiple cloning region	10–113
SP6 RNA polymerase promoter (–17 to +3)	124–143
SP6 RNA polymerase transcription initiation site	126
pUC/M13 Reverse Sequencing Primer binding site	161–177
lacZ start codon	165
lac operator	185–201
β-lactamase coding region	1322–2182
phage f1 region	2365–2820
lac operon sequences	2821–2981, 151–380
pUC/M13 Forward Sequencing Primer binding site	2941–2957
T7 RNA polymerase promoter (–17 to +3)	2984–3

Schematic representation of p GEMT-easy vector with the M13 forward and reverse primers located numerically according to coordinates. Multiple cloning region is shown between 10-113 bp (not labeled).

Appendix B: (Continued) pGEM-T Easy Vector MCS Map (Promega Biotech)



M13 Forward: 5’-GTTTTCCCAGTCACGAC-3’

M13 Reverse: 5’-CAGGAAACAGCTATGAC-3’

Appendix C: pGL4.23 Expression Vector System Map (Promega Biotech)

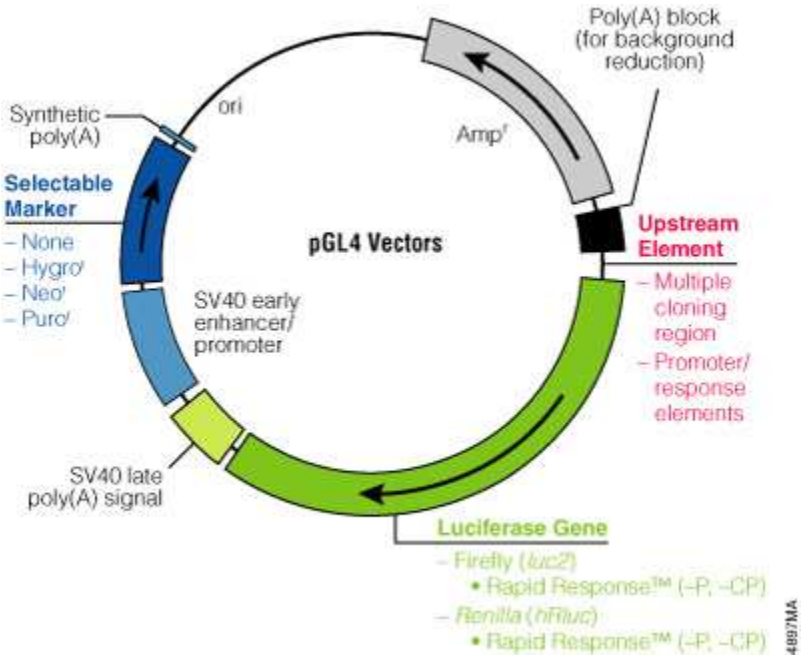
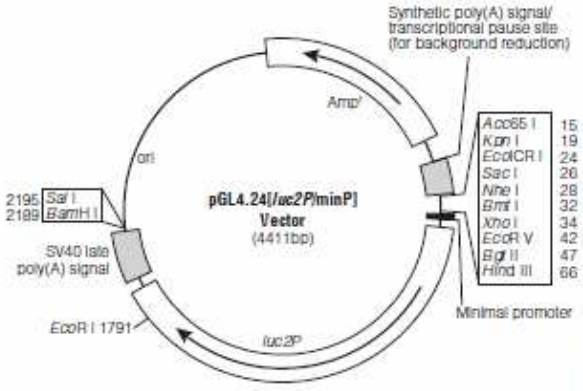


Figure 23. pGL4.23[*luc2*/minP] Vector circle map and sequence reference points.



Multiple cloning region	1–70
Minimal promoter	78–108
luc2P reporter gene	141–1916
SV40 late poly(A) signal	1956–2177
Reporter Vector primer 4 binding region	2245–2264
*Col*E1-derived plasmid replication origin	2502
Synthetic β-lactamase (Amp^r) coding region	3293–4153
Synthetic poly(A) signal/transcriptional pause site	4258–4411
Reporter Vector primer 3 binding region	4360–4379

Figure 24. pGL4.24[*luc2P*/minP] Vector circle map and sequence reference points.

Appendix C: (continued) pGL4.23 Expression Vector System MCS Map (Promega Biotech)

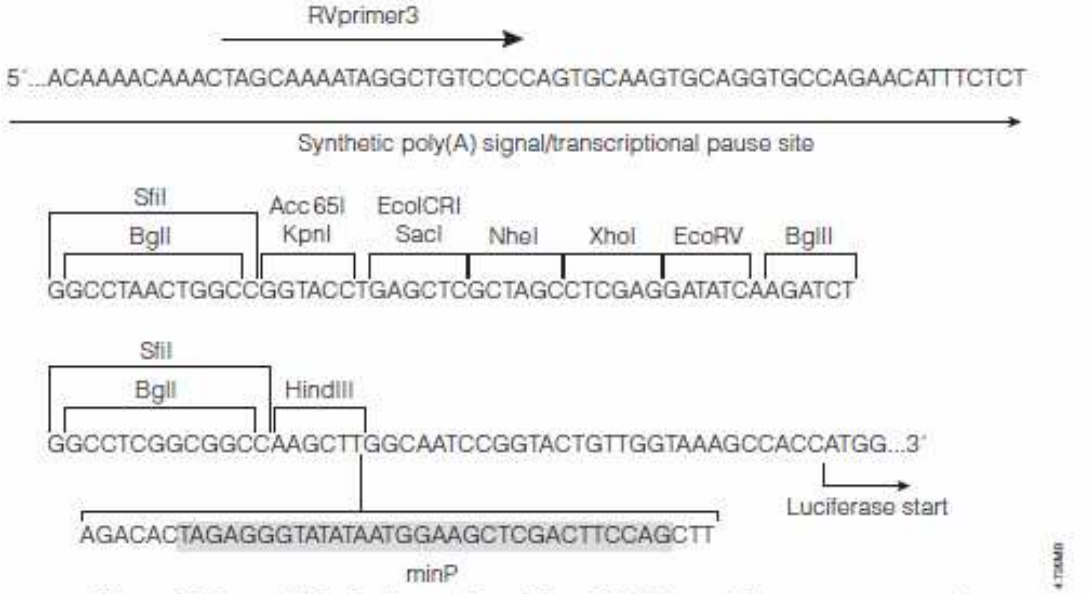


Figure 9. The multiple cloning region of the pGL4 Vectors. This map represents the promoterless pGL4 Vectors but also demonstrates the position and sequence of the minimal promoter, minP, found in the pGL4.23–pGL4.28 Vectors. The minP sequence is shaded in gray.

Appendix D: Restriction Enzyme (RE cut sites) found within ZFRYR3 by NEBcutter2.0, along with compatibilities of a selection of enzymes with pGL4.23.

Coordinates	Protein ID	Closest enzyme at 5' end	Closest enzyme at 3' end	Show flanking enzymes
compl(1478..1705)	-	MseI	EcoP15I	show
compl(548..745)	-	MwoI	HpyCH4V	show
compl(1344..1517)	-	NlaIV	SspI	show
912..1058	-	*Hpy166II	BsmAI	show
1058..1183	-	HpyCH4III	DraI	show
compl(1125..1244)	-	MnlI	DraI	show
438..551	-	DdeI	CviKI-1	show
792..896	-	HpyCH4V	*Hpy166II	show
152..244	-	MboII	Hpy188III	show
compl(302..391)	-	NdeI	HpyCH4V	show
339..425	-	CviKI-1	HpyCH4V	show
1688..1768	-	BccI	PleI	show
1301..1369	-	HpyAV	HpyAV	show
compl(1282..1344)	-	MseI	*BfuCI	show

Appendix D: (Continued)

pGL4.23[*luc2*/minP] Vector Restriction Enzyme Tables

The following restriction enzyme tables were constructed using DNASTAR® sequence analysis software. This information has not been verified by restriction digestion with each enzyme listed. The location given specifies the 5′ end of the cut DNA (the base to the right of the cut site). Vector sequences are also available in the GenBank® database (GenBank®/EMBL Accession Number DQ904455) and on the Internet at: **www.promega.com/vectors/.**

Restriction Enzymes That Cut the pGL4.23[*luc2*/minP] Vector Between 1 and 5 Times.

Enzyme	No. of Sites	Location
Acc65I	1	15
AccI	2	2068, 3748
AflIII	1	2317
AgeI	1	1273
Alw26I	1	4081
Alw44I	1	2631
AlwNI	1	2733
ApaI	1	172
AvaI	3	34, 688, 1354
AvaII	3	200, 1446, 3377
BalI	3	1123, 1648, 3473
BamHI	1	2061
BanII	2	26, 172
BbeI	1	1093
BbsI	2	200, 2146
BglI	2	9, 60
BglII	1	47
BlpI	1	1096
BsaBI	2	589, 2060
BsaMI	2	1880, 1973
BspHI	3	725, 1331, 3037
BspMI	1	4254
BsrGI	1	632
BssSI	2	1069, 2490
BstEII	1	3164
BstXI	1	3161
BstZI	3	1812, 1816, 3137
Bsu36I	1	3595
ClaI	2	2054,4182
DraI	3	2020, 3076, 3095
DraIII	1	1288
DrdI	2	1543, 2425
EarI	3	1450, 2201, 3457
EclHKI	1	3239
Eco47III	3	235, 346, 2193
EcoICRI	1	24
EcoRV	1	42
FseI	1	1818
FspI	1	964
HincII	3	1347, 1959, 2069
HindIII	1	66
HpaI	2	1347, 1959
Hsp92I	4	1090, 1538, 1733, 3776
KasI	1	1089
KpnI	1	19
NaeI	5	207, 1551, 1572, 1816, 2187
NarI	1	1090
NcoI	1	140
NgoMIV	5	205, 1549, 1570, 1814, 2185
NheI	1	28
NotI	1	3137
NspI	3	633, 888, 2321
PleI	5	1811, 2211, 2696, 3228, 3851
PshAI	1	2132
PstI	1	3161
PvuI	1	3609
PvuII	1	216
SacI	1	26
SacII	1	3633
SalI	1	2067
ScaI	4	307, 3719, 3811, 4189
SfiI	2	9, 60
SgrAI	1	205
SinI	3	200, 1446, 3377
SpeI	1	4068
SspI	2	4043, 4099
StyI	2	140, 755
XbaI	1	1799
XhoI	1	34

Appendix E: Detailed BLASTn search results

```
BLASTN 2.0MP-WashU [04-May-2006] [linux26-i686-ILP32F64 2006-05-
09T11:47:08]

Copyright (C) 1996-2006 Washington University, Saint Louis, Missouri
USA.
All Rights Reserved.

Reference:  Gish, W. (1996-2006) http://blast.wustl.edu

Notice:  this program and its default parameter settings are optimized
to find
nearly identical sequences rapidly.  To identify weak protein
similarities
encoded in nucleic acid, use BLASTX, TBLASTN or TBLASTX.
Query=  unnamed
        (1849 letters)
Database:  Danio_rerio.Zv8.56.dna.toplevel.fa
           11,624 sequences; 1,481,257,891 total letters.
WARNING:  Use of the hspsepSmax parameter should be considered with
long
          database sequences, to improve the biological relevance of
the HSP
          groups that are assembled and to improve the statistical
          discrimination of these groups from random background.
Searching....10....20....30....40....50....60....70....80....90....100%
done
WARNING:  hspmax=1000 was exceeded by 25 of the database
sequences,causing the associated cutoff score, S2, to be transiently
set as high as 16.
                                                                  Smallest
                                                                  Sum
                                                          High  Probability
Sequences producing High-scoring Segment Pairs:           Score  P(N)
N

20 dna:chromosome chromosome:Zv8:20:1:51884995:1          572  0.
4
14 dna:chromosome chromosome:Zv8:14:1:52930158:1          165  1.1e-172
11
7 dna:chromosome chromosome:Zv8:7:1:76918211:1            129  8.6e-164
8
1 dna:chromosome chromosome:Zv8:1:1:59305620:1            146  2.5e-160
12
21 dna:chromosome chromosome:Zv8:21:1:47572505:1          168  1.4e-152
11
9 dna:chromosome chromosome:Zv8:9:1:54736511:1            151  5.6e-152
9
13 dna:chromosome chromosome:Zv8:13:1:50748729:1          147  4.4e-151
8
5 dna:chromosome chromosome:Zv8:5:1:74451498:1            142  6.0e-151
9
2 dna:chromosome chromosome:Zv8:2:1:58009534:1            127  1.2e-150
11
```