Integrative characterization of the molecular evolution and functional divergence of teleost troponin C paralogs

by

Christine Elizabeth Genge
M.Sc., Queen’s University, 2010
B.Sc. (Hons), Queen’s University, 2007

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Biomedical Physiology & Kinesiology
Faculty of Science

© Christine Elizabeth Genge 2016
SIMON FRASER UNIVERSITY
Spring 2016

All rights reserved.
However, in accordance with the Copyright Act of Canada, this work may be reproduced, without authorization, under the conditions for Fair Dealing. Therefore, limited reproduction of this work for the purposes of private study, research, education, satire, parody, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.
Approval

Name: Christine Elizabeth Genge
Degree: Doctor of Philosophy
Title: Integrative characterization of the molecular evolution and functional divergence of teleost troponin C paralogs

Examining Committee: Chair: Dr. William Cupples
Dr. Glen Tibbits
Senior Supervisor
Professor

Dr. William Davidson
Supervisor
Professor

Dr. Thomas Claydon
Supervisor
Associate Professor

Dr. Jack Chen
Internal Examiner
Professor
Department of Molecular Biology and Biochemistry

Dr. Patricia Schulte
External Examiner
Professor
Department of Zoology
University of British Columbia

Date Defended/Approved: March 30, 2016
Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016
Abstract

Gene duplication results in extra copies of genes that can be sub-functionalized on structural and/or regulatory levels. Multiple paralogs are expressed in teleosts for troponin (Tn) components of the contractile unit (TnC, TnI and TnT), likely to maximize survival in different environmental conditions. The evolution of Tn subunits can be used as a model for understanding the variation in contractile function in ectotherms. The studies in this dissertation integrate evolutionary analysis with structural information to expand upon the knowledge of Tn function across phylogeny. Multiple parameters of cardiac structure and function were determined in vivo in the adult zebrafish, using high-resolution echocardiography, to accurately characterize the responses of this teleost model to both acute and chronic temperature perturbations. Cardiac output was modulated primarily by heart rate in response to acute temperature changes. With cold acclimation, a decreased E/A ratio suggests an increased reliance on atrial contraction for ventricular filling. The evolutionary history and sub-functionalization of the cardiac-specific TnC1 genes were characterized on both regulatory and structural levels. Three paralogs of TnC exist in fish, two of which are homologous to mammalian TnC1/cTnC. The TnC1 paralogs are likely the result of a tandem gene duplication that occurred in the common ancestor of the teleosts. In both zebrafish and trout hearts, TnC1 paralogs display temperature and chamber specific patterns in their usage of mRNA transcripts. While the zebrafish TnC1 paralogs have minimal variation in structure based on homology models, TnC1b has a higher Ca$^{2+}$ affinity relative to TnC1a as measured by isothermal titration calorimetry. Variation in the apparent affinity of zebrafish TnC1 paralogs for Ca$^{2+}$ results from dynamic conformational flexibility changes rather than from the direct interaction of site II with Ca$^{2+}$. Finally, the inter-related roles of regulatory and structural sub-functionalization that guide the co-evolution of interacting proteins in the Tn complex were explored. Transcriptional expression patterns predict various TnC/TnI complexes exist in the zebrafish heart with differential interaction strengths between the N-TnC and TnI switch region. Domain-specific divergent selection pressures and interaction energies suggest that substitutions in the TnI switch region are crucial to modifying TnI/TnC function to maintain cardiac contraction with temperature changes. Through these studies, the interacting proteins of the Tn complex have been established as an important model of the functional divergence of paralogs in the adaptive evolution of the teleost cardiac contractile element.

Keywords: teleost; genome duplication; sub-functionalization; functional divergence; troponin; cardiac
Acknowledgements

There are many people to thank for invaluable contributions through the duration of this degree. I was very fortunate to have an exceptional supportive supervisory committee. Firstly my appreciation goes to my senior supervisor, Dr Glen Tibbits, for providing ongoing guidance and support for my research, allowing me to focus on independent learning and collaborative skills. Thanks also goes to my committee members: Dr William Davidson, for his ongoing support and innumerable conversations to guide my scientific thinking process; and Dr Thomas Claydon for encouragement and open-mindedness as well as providing the qPCR equipment essential to two chapters of this thesis. I also thank Dr Jack Chen and Dr Patricia Schulte for serving as examiners.

Numerous coauthors on the manuscripts should be thanked both for sharing technical expertise as well as allowing for a much broader range of techniques to be used in the studies. Thanks particularly needs to go to Dr Charles Stevens, Kaveh Rayani and Lilian Lee for contributing technical skills in molecular dynamic simulations, ITC and echocardiography respectively which otherwise would not have been possible. I also have to thank Bruce Leighton for his assistance in designing and building the fish set-up used for the acclimation experiments, and Haruyo Kashihara for assisting in all the fish sampling. Both Bruce and Haruyo went above and beyond in helping to make sure these experiments went as smoothly as possible.

I have had the pleasure of working with and learning from many supportive colleagues as part of CMRL at SFU over the years. Particular thanks needs to go to Laura and Charlie for countless discussions about ideas and interpretations of data, as well as endless support and patience during more frustrating moments.

Finally I am extremely grateful for my family and friends who have always had my back and kept me on track. In writing this I realize how extensive this network is, and how important it has been to this process.
Table of Contents

Approval ............................................................................................................................. ii
Ethics Statement ............................................................................................................... iii
Abstract ............................................................................................................................. iv
Acknowledgements ........................................................................................................... v
Table of Contents .............................................................................................................. vi
List of Tables ..................................................................................................................... x
List of Figures ................................................................................................................... xi
List of Acronyms .............................................................................................................. xiii
Glossary .......................................................................................................................... xvi

Chapter 1. General Introduction .................................................................................. 1
  1.1. Excitation-contraction coupling ............................................................................. 2
    1.1.1. The sarcomere .............................................................................................. 3
  1.2. The troponin complex ........................................................................................... 4
    1.2.1. TnC .............................................................................................................. 7
    1.2.2. TnI ............................................................................................................. 8
    1.2.3. TnT .............................................................................................................. 9
    1.2.4. The Tn complex in ectothermic fish .......................................................... 10
    1.2.5. Factors affecting Ca²⁺ sensitivity .................................................................. 10
  1.3. Molecular evolution of Tn in fish ........................................................................... 16
    1.3.1. Gene duplication ....................................................................................... 16
    1.3.2. Protein evolution ........................................................................................ 19
    1.3.3. Sub-functionalization on a regulatory and structural scale ....................... 21
    1.3.4. Intermolecular coevolution ........................................................................ 22
    1.3.5. Current models of coevolution of protein-protein interactions ................. 24
    1.3.6. Evolution of the Tn family ......................................................................... 25
    1.3.7. Subfunctionalization in Tn family .............................................................. 31
  1.4. Objectives ............................................................................................................. 33
  1.5. References ............................................................................................................. 34

Chapter 2. Using functional divergence to characterize Tn paralogs ......................... 49
  2.1. Estimations of phylogenetic history .................................................................... 49
  2.2. Phylogenetic approach to understanding the fate of duplicates ....................... 55
  2.3. Quantification of variation in evolutionary rates ............................................... 57
  2.4. Detecting functional divergence beyond evolutionary rates ............................ 58
  2.5. Quantifying sub-functionalization of gene duplicates ...................................... 60
  2.6. Current limitations in the phylogenetic approach ............................................ 63
  2.7. References .......................................................................................................... 67

Chapter 3. Variation in cardiac contractility in fish ................................................. 72
  3.1. The fish heart ....................................................................................................... 73
    3.1.1. General morphology .................................................................................. 74
    3.1.2. Cardiomyocytes ....................................................................................... 77
3.1.3. Lifestyle demands influence morphology .................................................. 77
3.2. Filling patterns ........................................................................................................ 79
3.3. Force generation .................................................................................................... 80
3.4. Action potential ....................................................................................................... 82
3.5. Excitation-contraction coupling ............................................................................. 83
  3.5.1. Activation of Contraction ........................................................................... 84
  3.5.2. Relaxation of Contraction .......................................................................... 86
3.6. Variation in cardiac contraction with environmental perturbation ....................... 87
3.7. Summary ................................................................................................................ 90
3.8. References ............................................................................................................. 92

Chapter 4. Functional assessment of cardiac responses of adult zebrafish
(Danio rerio) to acute and chronic temperature change using high-resolution echocardiography ...................................................... 102
4.1. Abstract ................................................................................................................ 103
4.2. Introduction .......................................................................................................... 104
4.3. Methods ............................................................................................................... 106
  4.3.1. Animals and cold acclimation .................................................................. 106
  4.3.2. Zebrafish holding conditions .................................................................... 107
  4.3.3. Anaesthetic preparation .......................................................................... 107
  4.3.4. Echocardiography ................................................................................... 108
  4.3.5. Statistics .................................................................................................. 111
4.4. Results ................................................................................................................. 112
  4.4.1. Functional Analyses ................................................................................ 112
  4.4.2. Pulse wave Doppler measurements ........................................................ 114
  4.4.3. E/A values ............................................................................................... 115
  4.4.4. The Myocardial Performance Index (MPI). ............................................... 115
4.5. Discussion ............................................................................................................ 118
  4.5.1. Heart rate is primary cardiac function response to acute temperature changes .............................................................................. 119
  4.5.2. Zebrafish CO shows no response to cold acclimation .................................. 120
  4.5.3. Filling patterns do not change in response to acute temperature or acclimation .............................................................. 121
4.6. Summary .............................................................................................................. 123
4.7. Acknowledgements .............................................................................................. 123
4.8. References ........................................................................................................... 124

Chapter 5. Sub-functionalization of two distinct troponin C paralogs of the
teleost heart in a chamber and temperature dependent manner ..... 129
5.1. Abstract ................................................................................................................ 130
5.2. Introduction .......................................................................................................... 130
5.3. Methods ............................................................................................................... 133
  5.3.1. Sequence and phylogenetic analyses ..................................................... 133
  5.3.2. Animals and cold acclimation .................................................................. 134
  5.3.3. RNA analysis ........................................................................................... 135
5.4. Results ................................................................................................................. 137
  5.4.1. Troponin C sequence variation ............................................................... 137
Chapter 6. **Structural sub-functionalization of Ca\(^{2+}\) binding properties of zebrafish cardiac troponin C paralogs** ........................................159
   6.1. Abstract .................................................................................................................. 160
   6.2. Introduction .............................................................................................................. 161
   6.3. Experimental Procedures .......................................................................................... 164
       6.3.1. Homology Modelling and Equilibrium Molecular Dynamics
              Simulations ........................................................................................................... 164
       6.3.2. Free Energy Calculations .................................................................................... 165
       6.3.3. Umbrella Sampling & WHAM ......................................................................... 166
       6.3.4. Protein Expression & Purification ....................................................................... 166
       6.3.5. Melting Point Determination ............................................................................. 167
       6.3.6. Isothermal Titration Calorimetry ..................................................................... 167
       6.3.7. Functional divergence of amino acids ................................................................ 168
       6.3.8. Analysis of local frustration .............................................................................. 168
   6.4. Results ..................................................................................................................... 169
       6.4.1. Homology Models .............................................................................................. 169
       6.4.2. Free Energy Calculations .................................................................................... 174
       6.4.3. Melting Point determination .............................................................................. 176
       6.4.4. Isothermal Titration Calorimetry ..................................................................... 176
       6.4.5. Functionally divergent sites .............................................................................. 177
       6.4.6. Analysis of local frustration .............................................................................. 178
   6.5. Discussion ................................................................................................................ 180
   6.6. Summary .................................................................................................................. 185
   6.7. Acknowledgements ................................................................................................. 186
   6.8. References .............................................................................................................. 187
   6.9. Supplementary ....................................................................................................... 193

Chapter 7. **Functional divergence in teleost cardiac troponin paralogs guides variation in the interaction of TnI switch region with TnC** ............................................................ 201
   7.1. Abstract ................................................................................................................... 202
   7.2. Introduction ............................................................................................................ 202
   7.3. Methods .................................................................................................................. 208
       7.3.1. Collection of available TnI and TnC sequences .................................................. 208
       7.3.2. Phylogenetic analysis .................................................................................... 209
Chapter 7. Results

7.3.3. Zebrafish acclimation ................................................................. 210
7.3.4. Tissue-specific transcriptional expression .................................... 210
7.3.5. Molecular Dynamics (MD) ......................................................... 210
7.3.6. dN/dS ratios for selective pressures .......................................... 212
7.3.7. Functional divergence of amino acids ......................................... 212

7.4. Results ......................................................................................... 213
7.4.1. Phylogenetic analyses ............................................................... 213
7.4.2. Transcript expression patterns .................................................... 217
7.4.3. Equilibrium Molecular Dynamics and homology models ............ 218
7.4.4. Selection pressures .................................................................... 220
7.4.5. Functional divergence ............................................................... 222

7.5. Discussion ..................................................................................... 225
7.5.1. Differential expression patterns of TnI paralogs combined with structural variation in TnIswitch/TnC interaction suggest functional divergence between paralogs .............................................................. 225
7.5.2. Purifying selection occurs in all paralogs of TnI and TnC but domain-specific divergent selection patterns suggest TnI switch region variation distinguishes between TnI paralogs .................... 227
7.5.3. Type II functional divergent patterns in TnC increase variability in interaction with TnI switch region but do not influence interaction strength ................................................................................... 228
7.5.4. Functionally divergent residues in TnI switch region are not all interaction points with TnC but modify interaction strength ........................................................................................................... 229

7.6. Summary ....................................................................................... 231
7.7. Acknowledgements ........................................................................ 232
7.8. References ..................................................................................... 233
7.9. Supplementary .............................................................................. 234

Chapter 8. General Discussion ........................................................... 257
8.1. Aim 1: The nature of the variation in cardiac function in teleosts ........ 257
8.1.1. Variation in contractility is linked to mixed expression patterns of paralogs ........................................................................... 257
8.2. Aim 2: The context-specific sub-functionalization of the cardiac-specific TnC genes ................................................................. 261
8.2.1. Gene duplication in driving sub-functionalization ................................ 261
8.2.2. Inter-relation of regulatory sub-functionalization and structural sub-functionalization ........................................................................ 262
8.2.3. Structural sub-functionalization of TnC paralogs ................................ 265
8.3. Aim 3: Correlated evolutionary patterns in the interacting members of the Tn complex TnC and TnI ................................................................. 270
8.3.1. Protein-protein interactions ......................................................... 270
8.3.2. The Tn complex as a model of coevolution between subunits ....... 274
8.4. Future Directions ........................................................................... 277
8.5. References ..................................................................................... 280
List of Tables

Table 4-1 Parameters from echocardiographic B-mode of the zebrafish heart for WA vs CA fish at 18°C and 28°C ................................................................. 114
Table 4-2 Parameters from pulse wave Doppler and colour Doppler imaging of ZF heart ............................................................................................. 116
Table 5-1 Primers used for qRT-PCR analyses of zebrafish (ZF) and rainbow trout (RT) ............................................................................................. 137
Table 6-1 Average AB inter-helical angle over 5 replicated 1 µs simulations of Ca²⁺ bound and Ca²⁺ free TnC1a (cTnC) and TnC1b (ssTnC) ............ 171
Table 6-2 Average hydrophobic solvent accessible surface area (nm) over the final 50 ns of 5 replicated 100 ns simulations ........................................... 172
Table 6-3 Average number of hydrogen bonds in each structure over the final 100 ns of 5 replicated 1 µs simulations for Ca²⁺ bound and Ca²⁺ free models of TnC1a and TnC1b ................................................. 172
Table 6-4 Free energy of Ca²⁺ interaction calculated through Potential of Mean Force (kJ/mol) Errors are the standard deviation from 10,000 bootstraps .................................................................................. 175
Table 6-5 Thermodynamic Parameters derived from ITC for each paralog-temperature combination (n=3). For each parameter mutant-temperature conditions not having the same superscripted letter are significantly different p<0.05 ........................................................................................ 177
Table 6-6 Single Residue Level Frustration Index ........................................... 179
Table 7-1 Nomenclature for Tn paralogs ........................................................... 204
Table 7-2 Results of Clade model C (CM3) testing for divergent selection among codons between TnC paralogs .................................................. 221
Table 7-3 Results of Clade model C (CM3) testing for divergent selection among codons between TnI paralogs .................................................. 222
List of Figures

Figure 1-1  Schematic of the role of Troponin in Muscle Contraction. ............................... 6
Figure 1-2  Simplified schematic of relative timing of genome duplication events in vertebrates. ................................................................. 17
Figure 1-3  Potential fates of duplicate gene pairs with multiple regulatory regions according to the DDC model ..................................................... 19
Figure 3-1  Simplified morphology of a typical teleost heart. ............................................ 74
Figure 4-1  Longitudinal position of zebrafish. The ultrasound transducer beam (70 MHz) was positioned at about 85 degrees to ventral surface of the anesthetized zebrafish ....................................................... 108
Figure 4-2  B-mode image of a long-axis view of a WA zebrafish heart at 28°C................................................................. 110
Figure 4-3  Pulse wave Doppler image of ventricular inflow and outflow velocity ................................................................. 111
Figure 4-4  Ventricular functional parameters were calculated from B-mode long-axis plane (n = 10) ................................................................. 113
Figure 4-5  Ventricular inflow, made up of two components: early (E) filling peak velocity occurring during ventricle relaxation (E peak vel in mm/s) and the atrial (A) filling peak velocity resulting from active atrial contraction (A peak vel in mm/s), and ventricular outflow peak velocity (V outflow vel in mm/s). ................................................................. 117
Figure 5-1  Sequence comparison of representative TnC1a and TnC1b sequences from fish and human ................................................................. 139
Figure 5-2  Phylogenetic tree generated by the maximum likelihood comparison of TnC genes sequences across vertebrates demonstrating a third fish-specific paralog ................................................................. 141
Figure 5-3  Evolutionary history of TnC1a (cTnC) and TnC1b (ssTnC) amino acid sequences across fish ................................................................. 142
Figure 5-4  Quantitative real time PCR was used to determine tissue specific differences in relative mRNA levels of TnC paralogs in adult zebrafish at 28°C and adult rainbow trout at 5°C ................................................................. 144
Figure 5-5  Quantitative real time PCR was used to determine tissue specific differences in relative mRNA levels of TnC paralogs in adult zebrafish ................................................................. 145
| Figure 6-1 | Superimposed structures of the equilibrated representative structures after Daura clustering using backbone and beta carbon atoms over the duration of 100 ns equilibration trajectories. ........................................... 170 |
| Figure 6-2 | Sequence alignment of ZF TnC1a and TnC1b, highlighted to visualize the changes in secondary structure after equilibration at each temperature and energy minimization. .................................................. 170 |
| Figure 6-3 | Superimposition of snapshots of TnC1 demonstrated inter-helical angles ........................................................................................................... 171 |
| Figure 6-4 | Average root mean squared fluctuation per residue ............................................................................................................................... 173 |
| Figure 6-5 | Calcium coordination distances for each paralog-temperature combination. ....................................................................................................................... 173 |
| Figure 6-6 | PMF reaction Coordinate ......................................................................................................................................................................................... 174 |
| Figure 6-7 | WHAM derived Umbrella Potentials for each Temperature-paralog combination. ...................................................................................................................... 175 |
| Figure 6-8 | Type II functionally divergent sites in ZF TnC1 paralogs as detected by DIVERGE ........................................................................................................... 178 |
| Figure 6-9 | Variation in Helix C and D of ZF TnC1 paralogs ........................................................................................................................................................................ 179 |
| Figure 7-1 | Fish TnC1 divergence mapped onto human Tn complex structure (Takeda 2003 PDB 1J1D) ..................................................................................................................................................................................... 208 |
| Figure 7-2 | TnC maximum likelihood tree ......................................................................................................................................................................................... 215 |
| Figure 7-3 | Tnl maximum likelihood tree ......................................................................................................................................................................................... 216 |
| Figure 7-4 | Relative expression levels of TnI paralogs in zebrafish hearts ................................................................................................................................................. 217 |
| Figure 7-5 | Models and MM/PBSA interaction energies of the TnC/ TnI paralog combinations ................................................................................................................................. 219 |
| Figure 7-6 | Key sites undergoing divergent selection/functional divergence in the TnI switch/IR region (residue 106-143 in zebrafish Tnl1.1) ................................................................................................................................. 223 |
| Figure 7-7 | Interacting residues between TnI and TnC paralogs ................................................................................................................................................................. 224 |
### List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔΔCt</td>
<td>fold change in gene expression; change in threshold cycle</td>
</tr>
<tr>
<td>3R</td>
<td>teleost-specific whole genome duplication</td>
</tr>
<tr>
<td>AICc</td>
<td>corrected Akaike information criterion</td>
</tr>
<tr>
<td>ANOVA</td>
<td>analysis of variance</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>AV</td>
<td>atrioventricular</td>
</tr>
<tr>
<td>BEB</td>
<td>Bayes empirical Bayes</td>
</tr>
<tr>
<td>BI</td>
<td>Bayesian inference</td>
</tr>
<tr>
<td>BIC</td>
<td>Bayesian information criterion</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>BM2</td>
<td>branch model</td>
</tr>
<tr>
<td>CA</td>
<td>cold acclimated</td>
</tr>
<tr>
<td>Ca²⁺</td>
<td>calcium</td>
</tr>
<tr>
<td>CICR</td>
<td>Ca²⁺-induced calcium release</td>
</tr>
<tr>
<td>CM3</td>
<td>clade model C</td>
</tr>
<tr>
<td>CO</td>
<td>cardiac output</td>
</tr>
<tr>
<td>DDC</td>
<td>duplication-degeneration-complementation model</td>
</tr>
<tr>
<td>dN/dS</td>
<td>rate of non-synonymous/synonymous substitutions</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>E-C</td>
<td>excitation-contraction</td>
</tr>
<tr>
<td>E/A ratio</td>
<td>early to late filling ratio; early peak velocity / atrial peak velocity</td>
</tr>
<tr>
<td>EDV</td>
<td>end diastolic volume</td>
</tr>
<tr>
<td>EF</td>
<td>ejection fraction</td>
</tr>
<tr>
<td>ESTs</td>
<td>expressed sequence tags</td>
</tr>
<tr>
<td>ESV</td>
<td>end systolic volume</td>
</tr>
<tr>
<td>ET</td>
<td>ejection time</td>
</tr>
<tr>
<td>FD</td>
<td>functional divergence</td>
</tr>
<tr>
<td>Fₘₐₓ</td>
<td>maximal force of contraction</td>
</tr>
<tr>
<td>FS</td>
<td>fractional shortening</td>
</tr>
<tr>
<td>GROMACS</td>
<td>GROningen MAchine for Chemical Simulations</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>h-sasa</td>
<td>hydrophobic solvent accessible surface area</td>
</tr>
<tr>
<td>HR</td>
<td>heart rate</td>
</tr>
<tr>
<td>IR</td>
<td>inhibitory region of TnI</td>
</tr>
<tr>
<td>ITC</td>
<td>Isothermal Titration Calorimetry</td>
</tr>
<tr>
<td>IVCT</td>
<td>isovolumetric contraction time</td>
</tr>
<tr>
<td>IVRT</td>
<td>isovolumetric relaxation time</td>
</tr>
<tr>
<td>Kd</td>
<td>dissociation constant</td>
</tr>
<tr>
<td>LDA</td>
<td>length-dependent activation</td>
</tr>
<tr>
<td>LRT</td>
<td>likelihood ratio tests</td>
</tr>
<tr>
<td>MCMC</td>
<td>Markov chain Monte Carlo</td>
</tr>
<tr>
<td>MD</td>
<td>molecular dynamic simulations</td>
</tr>
<tr>
<td>MEGA</td>
<td>Molecular Evolutionary Genetics Analysis</td>
</tr>
<tr>
<td>MHC</td>
<td>myosin heavy chain</td>
</tr>
<tr>
<td>ML</td>
<td>maximum likelihood</td>
</tr>
<tr>
<td>MM/PBSA</td>
<td>Molecular mechanics Poisson–Boltzmann surface area</td>
</tr>
<tr>
<td>MPI</td>
<td>myocardial performance index</td>
</tr>
<tr>
<td>MRM</td>
<td>multiple reaction monitoring</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>MS222</td>
<td>Tricaine methanesulfonate</td>
</tr>
<tr>
<td>MSA</td>
<td>multiple sequence alignment</td>
</tr>
<tr>
<td>mtDNA</td>
<td>mitochondrial DNA</td>
</tr>
<tr>
<td>MUSCLE</td>
<td>Multiple Sequence Comparison by Log-Expectation</td>
</tr>
<tr>
<td>MYA</td>
<td>million years ago</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Centre for Biotechnology Information</td>
</tr>
<tr>
<td>NCX</td>
<td>Na⁺/Ca²⁺ exchanger</td>
</tr>
<tr>
<td>nDNA</td>
<td>nuclear DNA</td>
</tr>
<tr>
<td>NIQD</td>
<td>Asn2, Ile28, Gln29, and Asp30 in TnC1</td>
</tr>
<tr>
<td>NJ</td>
<td>neighbor-joining</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>PAML</td>
<td>Phylogenetic Analysis by Maximum Likelihood</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PKA</td>
<td>protein kinase A</td>
</tr>
<tr>
<td>PKC</td>
<td>protein kinase C</td>
</tr>
</tbody>
</table>
PLN phospholamban
PMF Potential of Mean Force
qRT-PCR quantitative reverse transcribed PCR
RMSD root mean square deviations
RMSF root mean square fluctuation
RSA relative solvent accessibility
s.e.m. standard error mean
SERCA SR Ca$^{2+}$-ATPase
SL sarcolemma
SR sarcoplasmic reticulum
SRLF single residue level frustration index
SV stroke volume
TF transcription factor
Tm tropomyosin
Tn troponin
TnC Ca$^{2+}$ binding troponin C
TnI inhibitory troponin I
TnISW TnI switch peptide
TnT tropomyosin-binding troponin T
WA warm acclimated
WCN weighted contact number
WGD whole genome duplication
$\Delta G$ Gibb’s free energy
$\Delta H$ change in enthalpy
$\Delta S$ change in entropy
Glossary

3R whole genome duplication event 320-250 million years ago prior to the divergence of the teleosts

acclimation remodeling to cope with longer term changes in temperature

Actinopterygii lineage originating from ray-fin fish; a subclass of bony fish

apo state of a protein without its binding cofactor i.e. TnC without Ca$^{2+}$ bound

bayesian inference probability that the tree is correct given the data and evolutionary model

branch model detects positive selection acting on particular lineages or branches of phylogeny

Ca$^{2+}$ affinity the Ca$^{2+}$ concentration at which 50% of Ca$^{2+}$ binding sites are saturated

Ca$^{2+}$ sensitivity the propensity for muscle contraction at a specific intracellular calcium concentration

cardiac compliance the ease of expanding or increasing volume of a chamber of a heart

cardiac contractility inherent ability for contraction of the heart during systole

cardiac output volume of blood pumped by the heart per minute; function of SV multiplied by HR

clade model detects divergent selection pressures between clades of a phylogenetic tree

co-evolution changes in at least two genes reciprocally affect each others evolution; within or between proteins, interacting residues compensate for affect of residue mutation by making a corresponding mutation

diastolic relaxation component of the cardiac cycle where ventricle refills

duplication-degeneration-complementation model (DDC) complementary degenerative changes in a pair of duplicate genes, such that the duplicates together retain the original functions of their single ancestor; degenerative mutations in regulatory elements can increase the probability of duplicate gene preservation

E-C coupling process connecting an electrical action potential and a mechanical muscle contraction

E/A ratio early peak velocity of filling ventricle / atrial or late peak velocity of filling the ventricle
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ectotherm</td>
<td>animal in which regulation of body temperature comes from external sources</td>
</tr>
<tr>
<td>EF-hand</td>
<td>An ion binding helix-loop-helix motif commonly found in Ca(^{2+}) and magnesium binding proteins</td>
</tr>
<tr>
<td>ejection fraction</td>
<td>volumetric fraction of blood pumped out of the ventricle with each cardiac cycle relative to the maximum volume in the heart after diastole</td>
</tr>
<tr>
<td>Frank-Starling</td>
<td>heart will eject a greater stroke volume at greater filling pressures</td>
</tr>
<tr>
<td>functional divergence</td>
<td>process by which duplicated genes shift in function from ancestral function; change in biophysical properties in the structure of a protein leading to functional diversity; correlated strongly with the change of evolutionary rate</td>
</tr>
<tr>
<td>type I</td>
<td>amino acid patterns are highly conserved in one gene but highly variable in the other gene</td>
</tr>
<tr>
<td>type II</td>
<td>amino acid patterns are highly conserved in both genes but have very divergent biochemical properties</td>
</tr>
<tr>
<td>homology</td>
<td>shared ancestry between proteins or genes in multiple species</td>
</tr>
<tr>
<td>in situ</td>
<td>within normal context</td>
</tr>
<tr>
<td>in vivo</td>
<td>within whole living organism</td>
</tr>
<tr>
<td>likelihood ratio tests</td>
<td>detects adaptive evolution; compares model to a null model where adaptive evolution is not allowed; significance in the LRT indicates the presence of sites evolving under significantly different selection pressures</td>
</tr>
<tr>
<td>maximum likelihood</td>
<td>probability of the data (such as a set of aligned sequences) being observed under a given phylogenetic tree and a specified model of evolution</td>
</tr>
<tr>
<td>multimeric</td>
<td>structure composed of several sub-units</td>
</tr>
<tr>
<td>myocardial plasticity</td>
<td>adaptability of the heart of an organism to changes in environmental conditions</td>
</tr>
<tr>
<td>neutral selection</td>
<td>dN=dS; no selective pressures where random mutations fix by chance</td>
</tr>
<tr>
<td>ortholog</td>
<td>genes in different species that evolved from a common ancestral gene</td>
</tr>
<tr>
<td>paralog</td>
<td>genes within a species that evolved from a common ancestral gene; separated by a duplication event</td>
</tr>
<tr>
<td>polyploidization</td>
<td>multiplication of whole chromosome content resulting in more than two paired homologous chromosomes</td>
</tr>
<tr>
<td>positive selection</td>
<td>dN&gt;dS; Darwinian selection pressure increasing frequency of beneficial alleles</td>
</tr>
</tbody>
</table>
protein stability  likelihood of protein to be in its native folded state; based on balance of thermodynamic forces

purifying selection  dN<dS; negative selective pressure; selective removal of deleterious alleles

Q_{10}  fold-change in activity per 10°C

regulatory sub-functionalization  context-dependent expression; duplicates expressed preferentially on a temporal or developmental scale, or in response to variable environmental factors; gain/loss of cis-regulatory elements between duplicates or modified post-transcriptional regulation

sarcomere  basic unit of striated muscle composed of thick and thin filaments that slide past each other during muscle contraction and relaxation

Sarcopterygii  lineage originating from lobe-fin fish; includes tetrapods

selective pressures  variations in evolutionary rates measured as the omega ratio of dN/dS = nonsynonomous (dN) and synonymous (dS) rates

structural sub-functionalization  changes within the coding sequence that modify the function of the protein products of gene duplicates

sub-functionalization  division of an ancestral function of a gene between the original and a duplicate

synteny  two or more genomic regions are derived from a single ancestral genomic region

systolic  contractile component of the cardiac cycle

tandem gene duplication  processes such as unequal crossing over during meiosis and resulting in closely spaced duplicate genes subject to reorganization

teleost  largest infraclass of ray-finned fish making up roughly half of extant vertebrate species; characterized by ability to protrude jaw

whole genome duplication  entire genome of an organism is doubled leading to polyploidization
Chapter 1.
General Introduction

Teleost fish exhibit a wide variation in thermal tolerance and capacity to acclimate to changing temperatures in order to cope with either long-term seasonal or acute temperature changes (155). For example, in ectothermic species where body temperature is susceptible to fluctuating environmental temperatures, the cardiac output must be regulated. Fish hearts must have intrinsic mechanisms or rely on post-translational modifications to protect against acute temperature changes when there is insufficient time for transcriptional regulation (161). Longer-term changes require remodeling of the cardiac tissue, which is often referred to as plasticity. Species-specific differences in fish cardiac function vary both in terms of tolerance and plasticity, which are each correlated with habitat conditions and thermal preferences (141). This capacity to cope with change has been linked to complexity introduced with teleost-specific whole gene duplication events (59).

Temperature greatly impacts cardiac contractility in ectothermic species. As an example, contributing factors to cardiac output such as heart rate and stroke volume demonstrably fluctuate with changing temperature in species such as salmonids (38). Cardiac remodeling may be influenced by changes in the susceptibility of the myofilament contractile apparatus to Ca\(^{2+}\) fluctuations. At lower temperatures, the Ca\(^{2+}\) sensitivity of the cardiac contractile unit decreases in both mammals (72) and fish (21) despite the critical role of myofilament Ca\(^{2+}\) sensitivity in regulating cardiac contractility. It has been postulated that ectotherms require heightened myofilament Ca\(^{2+}\) sensitivity to maintain cardiac function throughout select environmental changes.

In order to maintain cardiac output, critical components of the myofilament are modified in fish relative to mammals. The overall structure of the fish heart is unique, with
a simple parallel system of two contractile chambers responsible for circulation. While overall function varies, the underlying evolutionarily mechanism for producing these contractile changes can be identified. Some of the variability across fish species involves differential expression of paralogs, particularly, key proteins guiding the Ca²⁺ signaling of sarcomeric function. In this dissertation, the troponin (Tn) complex, as the key component that initiates contraction in response to Ca²⁺ transients, is the focus as a model of molecular evolution that produces functional divergence in a multimeric complex.

1.1. Excitation-contraction coupling

Calcium handling is the primary basis for regulation of cardiac contractility in all species, including fish. This is determined by the delivery and removal of Ca²⁺ to and from cardiac troponin C (cTnC), the thin filament protein that senses changes in cytosolic Ca²⁺ concentration and initiates the contractile process. The amplitude and kinetics of the cytosolic calcium transient modulate the kinetics and force of the contraction, and this in turn sets the limit for the contraction frequency. Membrane depolarization upon stimulation by an action potential causes an influx of Ca²⁺ through the L-type Ca²⁺ channel. An increase in local [Ca²⁺] causes opening of the ryanodine receptor in the sarcoplasmic reticulum via Ca²⁺-induced calcium release (CICR). This phenomenon alters the cytoplasmic [Ca²⁺] from submicromolar during diastole (relaxation) to micromolar during systole (contraction). The cytosolic concentration of Ca²⁺ increases from approximately 100 nM during diastole up to a maximum of ~1 µM during systole (10).

Mammalian cardiac muscle has intracellular Ca²⁺ stores in the sarcoplasmic reticulum (SR) in close proximity to T-tubules that provide the majority of the Ca²⁺ that binds to cTnC. In ectotherms, the activation of Ca²⁺ release from the SR and its contribution to contractile activation is controversial with some reports showing that the majority of Ca²⁺ is derived from extracellular space (39, 104, 147, 164) while others report a considerable contribution of the SR both to the activation of contraction (29, 84) and to Ca²⁺ removal from the cytosol during relaxation (84, 146, 147). Interestingly, Ca²⁺ handling by the SR has also been reported to differ between fish atrial and ventricular myocytes.
Chamber specific differences as well as species dependent variation are likely to account for part of this controversy on the role of the SR in the teleost heart. SR Ca\(^{2+}\)-ATPase (SERCA) and Na\(^+\)/Ca\(^{2+}\) exchanger (NCX) remove Ca\(^{2+}\) during relaxation and diastole but there is significant species-specific differences in the degree of dependence on each. The contributions of SERCA and NCX to relaxation roughly reflect the fraction of activation by the SR Ca\(^{2+}\) release or the Ca\(^{2+}\) influx from the sarcolemma (SL). Importantly, reverse mode NCX can contribute to activation of contraction. This is more likely to happen to a much larger extent in fish relative to mammals, and can even trigger SR Ca\(^{2+}\) release in trout atrial myocytes (83). Quantification of the contribution of each mechanism involved in the delivery and removal of Ca\(^{2+}\) to and from cTnC is necessary to estimate Ca\(^{2+}\) capacity of- and contribution to- the teleost cardiomyocyte. Overall, many components of the excitation-contraction machinery are conserved in both fish and mammals.

1.1.1. The sarcomere

The basic structure of the sarcomere is conserved across vertebrates. The major protein complexes that regulate cardiac contraction and relaxation include the thick filament and the thin filament, which form the characteristic striated pattern of cardiac and skeletal muscle. The contractile unit of muscle is the sarcomere, which runs from one Z-line to an adjacent Z-line. Myofibrils are composed of repeating units of sarcomeres. Within each sarcomere are parallel thick and thin filaments. The thick filament is primarily composed of myosin, the molecular motor protein of the sarcomere and titin. Myosin is composed of the myosin heavy chain (MHC) and four light chains each with alpha-helical structures that form parallel two-chain coiled structures. This forms the backbone of the thick filament with the N-terminal globular head (S1) at regular intervals. The S1 domain of myosin is the site of ATP hydrolysis that allows for binding of the thick filament to the thin filament during muscle contraction. Different isoforms of myosin, which hydrolyze ATP at different rates, are the determining factors in the different types of striated muscle. The tail of the myosin supercoil is rigid and aggregates with the tails of other myosin chains to form the central axis of the thick filament. The giant protein titin connects the myosin thick filament to the Z-disc, and contributes to passive stiffness of contractile muscle (62).
The proteins of the thin filament that regulate myosin motor activity include actin, tropomyosin (Tm), and the troponin (Tn) complex. The regulatory action of the troponin complex is transmitted through tropomyosin. Tropomyosin is an elongated protein that contains heptapeptide repeats each of which binds to seven actin monomers along each strand of the filament. This is an alpha-helical coiled dimer that lies along the groove of the actin filament, which covers the myosin binding sites at rest. The position of Tm on actin is regulated by Tn in response to Ca\textsuperscript{2+} binding.

The interaction between actin and myosin, catalyzed by the hydrolysis of ATP, is described as the cross-bridge cycle. Under the control of intracellular Ca\textsuperscript{2+} fluxes the thin and thick filaments slide past each other resulting in sarcomere shortening at the molecular level and contraction of muscle at the whole organ level. The sliding-filament theory of muscle contraction describes the role of the sarcomere in the transition between systole and diastole. During the cardiac cycle this involves three different states of the myofilaments: the blocked state, with complete steric blockage by tropomyosin of the actomyosin ATPase activity; the closed state, which involves weakly formed cross-bridges, and the open state which contains strongly bound actomyosin cross-bridges that are associated with sarcomeric contractions. Weak ionic interactions between myosin and actin occur independently of the presence or concentration of Ca\textsuperscript{2+}, which allows actomyosin cross-bridges to be prepared for activation just prior to systole in the closed state. Upon Ca\textsuperscript{2+} activation of the myofilaments, sarcomeric force generation is mediated by a strong hydrophobic interaction between myosin and actin and a subsequent isomerization of the myosin head by ATP release. These actin-myosin interactions are also seen as the rate limiting step for the myofilament relaxation. The specific biophysics of this cycle have been extensively reviewed (62, 152, 153) and the rest of this chapter will focus on the specifics relating to the Tn complex.

1.2. The troponin complex

Since Ca\textsuperscript{2+} handling is the primary basis for the regulation of cardiac contractility, the interactions involved in translating increases in Ca\textsuperscript{2+} to the sarcomere must be
discussed. In all species, including fish, the delivery and removal of Ca\(^{2+}\) to and from the troponin (Tn) complex guides the contractile process. Cardiac troponin (cTn) is made up of three proteins (cTnC, cTnI, cTnT). The cTn complex regulates actomyosin cross-bridge interactions in a Ca\(^{2+}\)-dependent manner in order to transition between contraction (systole) and relaxation (diastole). During systole, Ca\(^{2+}\) binds to cTnC activating the myofilament as the TnI switch arm binds to the N-terminal domain of cTnC. This allows cTnT to move tropomyosin across the thin filament causing cross-bridges to form between actin and myosin, which mediate contraction. During diastole, the release of Ca\(^{2+}\) from cTnC allows the regulatory arm of TnI to translocate to its inhibitory position on actin, blocking the actin-myosin cross-bridge activation. This results in sarcomeric relaxation. Through these interactions of the Tn complex, fluctuations in Ca\(^{2+}\) concentration thus signal the sarcomeric contraction and relaxation, both crucial to maintain appropriate cardiac function (Figure 1-1)(62, 153).
Figure 1-1  **Schematic of the role of Troponin in Muscle Contraction.**
The TnI switch region modulates the Ca$^{2+}$ sensitivity of TnC. Contraction of the cardiomyocyte occurs with the interaction of the thick and thin filament as mediated by Ca$^{2+}$ binding to TnC. When Ca$^{2+}$ binds to TnC, the TnI switch arm then binds to the N-terminal domain of TnC allowing TnT to move tropomyosin across the thin filament. This uncovers the myosin binding sites on actin thereby allowing cross-bridges to form between actin and myosin, ultimately initiating contraction. Structure of Tn generated from PDB 1J1D (159).
1.2.1. TnC

TnC is a small (18 kDa), highly conserved protein with two paralogs in mammals. cTnC (also referred to as TnC1), expressed in cardiac and slow skeletal myocytes, is made up of 161 amino acids while fsTnC (also referred to as TnC2) is expressed in fast skeletal myocytes and is made up of 162 amino acids. TnC is the Ca\textsuperscript{2+}-binding subunit of the troponin complex and consists of nine helices composing two globular domains connected by a short linker. Each domain has two Ca\textsuperscript{2+} binding EF-hand motifs (35, 51). The primary function of TnC is to bind Ca\textsuperscript{2+} and this is accomplished through the EF-hands binding Ca\textsuperscript{2+} through coordination of six charged residues in a twelve residue loop. The key residues located in a pentagonal bipyramidal coordinating structure are a hallmark of EF-hands (51). In mammalian fast skeletal TnC (fsTnC), all four EF-hand sites bind Ca\textsuperscript{2+} with affinities that are physiologically relevant. Cardiac TnC possesses only one low affinity Ca\textsuperscript{2+} binding site (site II binding affinity of 1 x 10\textsuperscript{5} M\textsuperscript{-1}) as site I has been rendered non-functional by amino acid substitutions that preclude Ca\textsuperscript{2+} coordination over the range of physiologically relevant Ca\textsuperscript{2+} concentrations (67, 110, 156, 162). The C-terminal EF-hands are high-affinity Ca\textsuperscript{2+}/Mg\textsuperscript{2+} sites (site III/IV binding affinities of 1x10\textsuperscript{7} M) (82) and anchor TnC to the troponin complex. These sites have a Ca\textsuperscript{2+} dissociation rate too slow to be responsible for contraction and relaxation (120) and are probably 100% saturated with Ca\textsuperscript{2+} over the range of physiological Ca\textsuperscript{2+} concentrations (51).

During the apo or unbound state, cytosolic Ca\textsuperscript{2+} concentrations are low (~100 nM) and the hydrophobic residues involved in the interaction with the regulatory switch arm of TnI remain inaccessible. The binding of Ca\textsuperscript{2+} to the lower affinity site II of cTnC results in a conformational change (79) which transitions TnC to an open state and exposes a hydrophobic patch on the surface of the N-terminal domain. This hydrophobic patch binds to the switch region of TnI (residues 147-163 in human cTnI), which pulls TnI away from its inhibitory position on the actin filament. A wide range of predicted structural changes occur with the binding of calcium to site II (49, 110), including a decrease in flexibility of the backbone of TnC and a more rigid loop in site I (156). In some ways these are priming events for binding of the regulatory arm of cTnI, since the activation energy may not be achieved in order to fully open the hydrophobic pocket of TnC (148). The flexible D-E linker
of TnC connecting the D and E helix is also a critical region. The transition between systole and diastole changes the regions involved in interactions with the inhibitory region of Tnl. Specifically residues E85 and D86 have been shown to be required for proper transitions that guide actomyosin activity (100). These interactions initiate and regulate force production of the contractile element, which emphasizes the role of cTnC as a key regulator of the myofilament Ca$^{2+}$ sensitivity.

1.2.2. Tnl

Troponin I (Tnl) is the inhibitory subunit of the Tn complex responsible for the inhibition of actomyosin ATPase activity. In mammals there are three isoforms, which are identified by independent tissue-specific expression patterns: slow skeletal (ssTnl, Tnl1, Tnnl1), fast skeletal (fsTnl, Tnl2, Tnnl2) and cardiac (cTnl, Tnl3, Tnnl3). In developing hearts, ssTnl is expressed and expression transitions to the adult cardiac isoform shortly after birth.

The first domain of Tnl consists of the N-terminal range of approximately 30 amino acids that interacts with the NH$_2$ lobe of cTnC that is present in mammalian cTnl but not in fish (126). The two protein kinase A (PKA) phosphorylation sites in this region can cause the N-peptide to be released from the strong interactions with cTnC resulting in a decrease in the Ca$^{2+}$ sensitivity of cTnC which promotes increased cross-bridge cycling (88). This region is important in the adrenergic control of Ca$^{2+}$ fluxes linked to increased heart rate, independently of Ca$^{2+}$ levels. The first alpha-helix of Tnl (residues 43-79 in human cTnl) binds the C-terminal end of cTnC. The C-terminal end of H1 (residues 66-79), a flexible linker (residues 80-89) and the second alpha-helix (residues 90-135) wrap around the H2 helix of cTnT (residue 226-271) and form a hydrophobic core. Many of these interactions between the H2 helix of cTnl and the H2 helix of cTnT are well conserved across species (41) and this region of cTnl and cTnT is defined as the IT arm in the Tn complex. The C-terminal end of cTnC is also involved with interactions in this domain making the IT arm critical in the overall conformation of the cTn complex.
cTnI binds to actin via two regions, a highly basic inhibitory peptide (residue 137-210 in human cTnI) and a second actin binding region. These regions flank the switch region, which is the critical point of contact with cTnC during Ca\(^{2+}\) binding. The C-terminal end is a mobile domain. These domains together are considered to be the regulatory domain of cTnI.

The inhibitory region (IR) plays a central role in the blocked state of the myofilament; it prevents actomyosin activity during diastole. Residues L145-R149 are required for activation of the myofilaments via binding of the IR to D-E linker of TnC. These residues are less important for inhibition of cross-bridge cycling via actin binding than for binding of the IR to TnC (99). TnC and actin appear to compete for access to this region.

The switch domain of TnI, a highly mobile region, is somewhat variable in length. It begins with amphiphilic helix 3 (residues 150-159) and ends with helix 4 (residues 164-188). During diastole its primary function is to provide a second actin and Tm binding site via a weak interaction (119). The movement of the switch arm between systole and diastole may play a role in altering the conformational state of the IR (41, 42). The translocation of the regulatory arm of TnI transitions the IR region from a beta-hairpin to an extended alpha-helical structure (32, 159), releasing Tm from its blocked position and uncovering the myosin binding sites on actin to allow for ATPase activity and contraction (80, 106, 130).

1.2.3. TnT

Troponin T is the subunit of Tn responsible for direct interactions with tropomyosin. There are three paralogs of TnT in mammals: muscle specific fast skeletal (TnnT3), cardiac (TnnT2) and slow skeletal (TnnT1). These paralogs are differentially regulated during embryonic and postnatal development as well as during physiological and pathological adaptations (18). Multiple alternative splice variants are expressed from each of the TnT genes adding further structural and functional variation. In the mammalian heart, cTnT splice variants are post-transcriptionally regulated temporally during embryonic and postnatal heart development, independent of functional demands (144).
The TnT protein ranges from 250-305 amino acids with molecular weights ranging from 31 kDa to 36 kDa. The large variation in size is due to the variable length of the N-terminal region. TnT is composed of two domains: the first is the extended N-terminal T1 domain (residues 1-158) and the second is the globular C-terminal T2 domain (residues 159-259 in human TnI1). This N-terminal region is hypervariable (133) with variable lengths and sequence composition thought to modulate the conformation and function of the TnT core structure (11). A highly conserved region of the N-terminal T1 domain interacts with the C-terminal domain of Tm (90). The T2 domain interacts both with the C-terminal end of TnC and the N-terminus of TnI as well as F-actin (144). TnT is critical in thin filament function via its role in connecting the signaling activities of TnC and TnI to the position of Tm in the actin groove, as well as anchoring the Tn complex to the thin filament.

1.2.4. The Tn complex in ectothermic fish

Much of the research regarding the Tn complex has been done in mammalian models. However, ectotherms are faced with multiple challenges to maintain effective contraction. Substantial variation between fish and mammals occurs in the Tn complex. Fish and mammalian myofilament Ca$^{2+}$ sensitivity is similar at their respective physiological temperatures, however, at comparable temperatures Ca$^{2+}$ sensitivity appears to be increased in fish (22, 74). Even the highly conserved TnC has particular modifications to deal with unique environmental challenges faced by fish. In order to understand the variation within teleost Tn, it is necessary to first understand the factors that affect Ca$^{2+}$ sensitivity. Many of these are environmental considerations important for the survival of ectothermic species.

1.2.5. Factors affecting Ca$^{2+}$ sensitivity

The Tn protein complex has a complex functional landscape that is not limited to Ca$^{2+}$ binding. Modification of the function of Tn can encompass a range of protein interactions that are not limited to changes in TnC Ca$^{2+}$ affinity. The ability of the regulatory domain of TnC to bind Ca$^{2+}$ is highly dependent on its interaction with other myofilament
proteins (120). These interactions are influenced by many external factors, some of which affect affinity, the direct energy of Ca\(^{2+}\) binding, whereas others influence sensitivity, which includes conformational changes that modify Ca\(^{2+}\) binding signal transduction to the thin filament. All of these factors have important implications for the function of the Tn complex and in turn cardiac contractile function in ectotherms.

Ca\(^{2+}\) sensitivity in cardiac myocytes is influenced much more by sarcomere length than skeletal myocytes (70). The contractile element and thus cTnC has an increased sensitivity for Ca\(^{2+}\) at longer sarcomeric lengths (7, 28, 114). This contributes to the Frank-Starling mechanism: increases in end diastolic volume translate into increased sarcomere lengths leading to increases in stroke volume (65). The twitch force is directly proportional to systolic sarcomere length (28, 153). Simple inter-filament spacing that causes the Frank-Starling phenomenon has now been questioned and length-modulated regulation of thin filament activation state is considered to be increasingly important (28). The increase in twitch force development with increase in sarcomere length has been linked with Ca\(^{2+}\) responsiveness of the cardiac contractile unit (28). Myofilament activation is highly cooperative as sharp increases in force development occur over a narrow range of Ca\(^{2+}\) concentration (30, 95, 98, 153). This may be due to several factors including: enhanced probability of Ca\(^{2+}\) binding to TnC when neighbouring TnC sites are occupied by Ca\(^{2+}\), cooperative feedback between troponin sites in activating the thin filament, promotion of further cross-bridge formation by near-neighbour active cycling, or increased Ca\(^{2+}\) binding affinity in proportion to the number of force generating cross-bridges (57, 103, 137). Stretch-induced structural rearrangement of the contractile proteins of the thin filament may be mediated by titin strain (1). The inhibitory region of cTnI has also been proposed to play a prominent role in the length transducing process of the sarcomere. A threonine residue at position 144 in mammalian cTnI was found to govern this property via protein kinase C (PKC) phosphorylation (158).

Fish, in particular, have stroke volumes that are greatly dependent on the amount of preload placed on the ventricle (23, 46, 132). Many active fish exhibit a broad dynamic range in activation of the myofilaments. In fish, twitch force has also been shown to increase in proportion to stretch, both in permeabilized cells and skinned myofibrils (zebrafish - (36); rainbow trout - (131)). The myofilament length dependency measured is
similar to that found in the mammalian sarcomere (115), as well as resting sarcomere length (145). This suggests that the length-dependent activation (LDA) is conserved across species.

Temperature has a large impact on the Ca\(^{2+}\) sensitivity of the entire contractile element in cardiac but not skeletal muscle (73). For all species, as temperature decreases, the Ca\(^{2+}\) concentration required to generate an equal amount of force increases. In cold water fish the cardioplegic influence of low temperature on myofibrillar Ca\(^{2+}\) sensitivity must be overcome by either greater Ca\(^{2+}\) delivery/removal or compensatory differences in Ca\(^{2+}\) sensitivity of the contractile apparatus. Higher Ca\(^{2+}\) sensitivity results in slower unloading of Ca\(^{2+}\) with lower temperatures, which goes along with decreased maximum heart rate (HR). The decrease in maximum HR is a consequence of lengthening the duration of the action potential. Greater Ca\(^{2+}\) fluxes could be produced in several ways, including alterations in the number of Ca\(^{2+}\) transporting proteins, enhancing their activity, or expression of functionally divergent paralogs of the Ca\(^{2+}\) handling proteins. Replacement of cTnC with fsTnC in mammalian cardiomyocytes relieves the desensitizing effect of low temperature on contractile strength (72). Nuclear magnetic resonance (NMR) studies reveal that trout cTnC has a similar conformation to mammalian cTnC at their respective physiological temperatures, which suggests that the teleost cTnC structure may be optimized for function at lower environmental temperatures (14). The temperature sensitive properties of cTnC are linked to specific amino acids shown to be responsible for the increased Ca\(^{2+}\) sensitivity found in trout cTnC relative to mammalian cTnC, including Asn2, Ile28, Gln29, and Asp30 (NIQD). These residues are hypothesized to allosterically affect the ability of site II to bind Ca\(^{2+}\) (58), allowing Ca\(^{2+}\) binding at lower temperatures than would otherwise be possible (56). These residues also restore some function to site I via allosteric effects (53).

A shift in pH is observed with shifts in temperature, typically by -0.16 to -0.19 pH/10 °C in poikilothermic fish (21). This relationship keeps the relative alkalinity ([OH\(^{-}\)]/[H\(^{+}\)]) constant. Modification of intracellular pH in response to temperature shifts, termed alpha-stat regulation (136), is thought to maintain cardiac myofilament function in ectotherms by keeping fractional dissociation of alpha-histidine imidazole groups approximately constant (21). The alpha-stat regulation strategy in poikilotherms can help compensate for acute
temperature effects. The stable protonation state of histidine imidazole groups rather than the pH value itself is conserved across temperatures. A small change in pH will result in a relatively large change in protonation state within physiological pH range, therefore the number of histidine residues may be an important determinant of the pH dependence of proteins (136).

Decreases in pH lead to decreased Ca\(^{2+}\) sensitivity associated with the rapid drop in the force production of cardiac muscle contraction due to acidosis (12, 40, 128). A higher Ca\(^{2+}\) concentration is needed to activate myofibrils at lower pH values (12) and maximal force produced is decreased at lower pH values (123). Lower pH results in a decrease in Ca\(^{2+}\) sensitivity of myofibrillar ATPase in the physiological ranges of both fish and mammals (22). Physiological pH changes only partially compensate for temperature-dependent desensitization thus higher myofibrillar Ca\(^{2+}\) sensitivity may be necessary for function at decreased temperatures. In the Tn complex, TnI has a residue identified as a pH sensitive histidine button, which is present in the slow skeletal TnI1 but not in cardiac TnI3 (8, 109). This results in the cardiac form of Tn being more susceptible to changes in pH, notably in the form of acidosis-induced contractile failure. TnC in intact myofibrils shows a greater effect of pH than TnC in isolation, suggesting the presence of other proteins are critical in pH responsiveness (154). However, TnC Ca\(^{2+}\) affinity is also directly affected by pH (56, 66, 128). A large reduction has been measured of Ca\(^{2+}\) binding directly to the functional site of canine cTnC at pH 6.2 relative to pH 7.0 (12), indicating the direct effect of acidosis. Proton exchange has since been shown to be important in the mechanism of Ca\(^{2+}\) coordination, likely a determinant of pH sensitivity for TnC (150).

The influence of thin filament proteins that interact with TnC play a large role in Ca\(^{2+}\) sensitivity (82) One physiologically important modulator of TnC Ca\(^{2+}\) affinity is the TnC-TnI interaction (82, 92). TnI inhibits the formation of actin-myosin cross-bridges during diastole and influences the ability of TnC to bind to Ca\(^{2+}\) during systole by directly increasing the size of the hydrophobic patch. The presence of TnI switch arm increases the size of the TnC hydrophobic patch (49, 110). The N-terminal region of TnI binding to TnC also increases the Ca\(^{2+}\) sensitivity of the sarcomere (88). In mammalian Tn, PKA-mediated phosphorylation of two serine residues in the N-terminal extension of cTnI modulates the interaction between cTnI and cTnC (111). Phosphorylation of these two
sites weakens the interaction between the N-terminal domain of cTnI and the N-terminal domain of TnC (78, 85), destabilizing the Ca$^{2+}$ bound state of cTnC. This decreases the Ca$^{2+}$ sensitivity of cTnC and allows for an increased rate of myocyte contraction with a higher rate of off-loading of Ca$^{2+}$ from the contractile element (157). This beta-adrenergic signaling through PKA is an important way that the heart modulates inotropic function during stress in mammals.

The absence of this N-terminal extension in teleost TnI may contribute to the enhanced myofilament sensitivity to Ca$^{2+}$ seen in ectothermic teleosts (reviewed in (143)). However, other phosphorylation sites for protein kinases A and C may be present in fish TnI that, when introduced in mammalian TnI, reduce the Ca$^{2+}$ sensitivity of myofilaments (97). PKC has also been coupled with regulation of cardiac contraction in mammals with known targets in TnI3 including S43/45 (68, 116, 169). PKC phosphorylation of S43/45 results in a decrease in maximum Ca$^{2+}$ activated force and cross-bridge cycling rate (15, 135). Phosphorylation of T144 decreases myofilament sliding velocity and may contribute to the relaxation effects observed with S23/24 phosphorylation (15, 170). PKC and PKA have somewhat opposing effects, but combined they produce a significant level of fine-tuning required for proper myofilament function. In addition, the pH responsive histidine at residue 164 in the TnI switch arm provides another example of the effect of TnI beyond pH responsiveness (8, 109). This histidine replacement by an alanine in mammals is proposed to be responsible for the increased myofilament Ca$^{2+}$ sensitivity, and decreased rate of myofilament relaxation relative to ectothermic fish (126).

Other interactions within the sarcomere play an important role in TnC Ca$^{2+}$ sensitivity. By facilitating the dissociation of TnI from actin, myosin can indirectly cause a transient increase in Ca$^{2+}$ sensitivity (9, 26). The myosin-actin interaction itself may cause structural changes in actin to move Tm and increase the probability of cTnC-cTnI binding (26). Full activation of the thin filament itself requires both Ca$^{2+}$ and rigor of myosin S1 (77).

Ectotherms face significant challenges in dealing with the factors that affect Ca$^{2+}$ sensitivity. Inter-relation of all these factors highlights the difficulties of looking at temperature changes independently when focusing on ectotherms. Simply examining TnC
to understand changes in Ca$^{2+}$ sensing is often insufficient since many events involving other proteins have repercussions on the entire contractile element. Another factor not discussed is the impact of multiple paralogs in the contractile function. Gene duplication events are thought to be the driving force in evolutionary novelty that plays a critical role in the maintenance of cardiac function over environmental extremes. The variation in paralog composition may be indicative of the mechanism of molecular evolution to cope with a variety of physiological challenges. The focus of the next section will be on the basis for how this variation in paralog expression and function evolved in ectothermic fish.
1.3. Molecular evolution of Tn in fish

Multiple copies of Tn subunits appear in all vertebrate genomes. Determining how homologous genes compare to each other is important for understanding the degree of relatedness. Orthologous genes appear between species, and should be compared to each other for meaningful phylogenetic conclusions. This becomes more difficult with polyploidization associated with the teleost-specific 3R genome duplication event. Within species, paralogs exist as extra gene copies while ohnologs are specific as products of whole genome duplication events. When using fish as a model it is important to consider that paralogs and ohnologs could create differential cardiac contractile properties. However, in order to appreciate how they are retained and their regulatory and/or structural function, the evolution of duplicates on both a gene and a protein level will be discussed.

1.3.1. Gene duplication

Duplications can affect individual genes, stretches of several genes, whole chromosomes or even whole genomes. These events are considered to be the major evolutionary source of new protein functions (45, 89, 121, 122). Tandem duplicates can arise from processes such as unequal crossing over during meiosis and usually result in closely spaced duplicate genes, that are subject to reorganization. Whole genome duplication (WGD) events describe when the entire genome of an organism is doubled leading to polyploidization. The “one-two-four” model is generally used to describe the two WGD events (1R and 2R) that occurred early in vertebrate evolution. A third WGD event (3R) occurred prior to the divergence of the teleosts (59) (Figure 1-2). The conserved synteny (gene order on chromosomes) of Hox clusters of genes in teleost fish has provided strong evidence for this 3R event (5). The timing of this duplication has been estimated to have occurred between 320-250 million years ago (20, 163), consistent with the timing of teleost radiation. More recent genome duplications have also taken place within teleost lineages, specifically salmonids and cyprinids.
While the one-two-four rule coupled with a fish-specific WGD event implies that a single-copy gene in basal deuterostomes should exist as four copies in mammals and eight copies in teleosts, duplicates are often differentially retained across lineages (59). If both duplicates are identical, one is usually lost due to genetic drift with accumulation of deleterious mutations in one of the duplicate genes. The fate of the duplicate pair is thought to depend on how selection impacts the subsequent redundancy in function. Redundant genes also provide possibilities for mutations to accumulate leading to evolutionary novelty without adversely affecting fitness. Many isoforms result from duplication events, either via WGDs or single-gene duplications. While some copies have neo-functionalized to take on novel roles with independent functions, others are the result of sub-functionalization and have partitioned ancestral functions (138).

The duplication of a gene results in an additional copy that is theoretically free from selective pressure. This allows the new copy of the gene to mutate without deleterious consequence to the organism. This freedom from consequences allows for the mutation
of novel genes that could increase the fitness of the organism or code for a new function. Under the original model of WGD, neo-functionalization was the primary mechanism for retention of duplicated genes. However, neo-functionalization is typically associated with positive selective pressures, while sub-functionalization is usually a purely neutral process (101, 138).

The duplication-degeneration-complementation model (DDC) suggests that degenerative mutations in regulatory elements can increase the probability of duplicate gene preservation (Figure 1-3). The usual mechanism of duplicate gene preservation is done through partitioning of ancestral expression patterns rather than modification of biochemical functions (44). For example, in zebrafish, many gene pairs resulting from WGD show evidence of changes in regulatory control and novel expression domains (94) spatially in tissue-specific patterns, temporally regarding developmental stage patterns, or in response to environmental perturbations such as temperature. Expression studies, especially in zebrafish, often do not include adult stages in analysis but often expression patterns of duplicate gene pairs become more distinct later in development (94).

The DDC has been expanded to include progressing with time to modification of coding sequences. This process occurs by extending the amount of time during which mutations can occur in coding sequences since the duplicates can be stabilized in the genome without detrimental dosage effects. Proportionately small changes in coding sequences can have a large effect on morphology if they occur in genes that have a prominent regulatory or structural role. Protein structure itself may play a role in amplifying the importance of positive and negative selection due to the fitness costs of changes in folding patterns in tertiary structure (174).
Figure 1-3  Potential fates of duplicate gene pairs with multiple regulatory regions according to the DDC model.
The small boxes denote regulatory elements with unique functions, and the large boxes denote transcribed regions of the gene. Solid boxes denote intact regions of a gene, while open boxes denote null mutations, and grey boxes denote the evolution of a new function. Note – these are not always discrete events. Mutations causing sub-functionalization in the regulation of gene expression can lead to further structural sub-functionalization on the level of the coding sequence modulating protein function, or even neo-functionalization. Visual depiction of concepts described in (44).

1.3.2.  Protein evolution

There are several major determinants of protein evolutionary rates (reviewed in (112)). The degree to which the protein is essential is correlated with slower evolutionary rates (81, 93, 125, 165). Typically protein evolution is conservative. When amino acid changes occur they are usually conservative in nature and may not have a functional consequence. Non-synonymous rates in particular are thought to be guided by the tolerance of the 3D structure of the protein for amino acids substitutions. Most of the time structural changes are even more conservative than sequence changes such that sequences can be very divergent while the structure is still maintained (105).

The stability of the native state of a protein is one of the major determinants of protein evolution (13, 31, 37, 71). The primary structural constraint on amino acid substitution is solvent accessibility. The solvent inaccessible cores of proteins are closely
packed, more hydrophobic and more conserved than surface regions (25). The hydrophobic effect contributes greatly to the stability of the folded state of the protein (174). Other factors involved with the stability of a protein include hydrogen bonds within alpha-helices and beta-sheets or through side chains. These bonds make comparable contributions to protein stability through increased packing density and stronger van der Waals interactions in the core of the protein. Buried side chains are among the most conserved amino acid residues in proteins, seen as a recurring pattern across families of proteins (174). Due to these structural constraints, residues within a protein do not evolve independently. Local structural environment or intramolecular constraints influence evolutionary rates. Active sites/functional residues are under a greater pressure to be conserved since they are crucial to protein function (61). A higher degree of residue conservation is also seen in residues in close proximity to the active site (16).

To ensure proper protein folding patterns, evolutionary selection against the detrimental effects of protein aggregation must occur (33, 34, 142). This suggests that protein stability can be used as part of the estimation of selection pressure (25). This is more of an issue for highly expressed proteins. Evolutionary rates are correlated with expression level, with highly and ubiquitously expressed proteins evolving at a slower rate (174). Changes in constraints in short linear motifs may lead to regulatory changes and therefore produce functional divergence. This has been hypothesized to be the result of more functionally conserved proteins being more conserved, which minimizes the risk of compromised function (33). A more highly expressed protein will have higher constraints to avoid misfolding and will show preferential usage of codons that are less likely to induce translation errors (174).

Along with improper protein folding, other biophysical mechanisms that should be considered include improper protein interactions. This is reflected in selection pressures being influenced by overall protein stability that ultimately acts on interactions between sites of interacting proteins (172), since interfacing sites do not evolve independently. Evolutionary rates are related to the number of protein-protein interactions, which can be related to their proximity to the active site in a similar way to intermolecular constraints (61). The function of a protein may not be independent, and interactions with functional partners to maintain favourable and avoid unfavourable interactions should also be
considered. Analyses on whole proteins may underestimate the level of functional divergence after gene duplication. Proteins have a modular structure typically composed of evolutionarily conserved functional units contributing to various catalytic and interaction functions (19). These domains are interspaced with variable domains largely contributing to their structure (96). Domains that are involved in interactions will thus be subject to further constraints.

1.3.3. **Sub-functionalization on a regulatory and structural scale**

Gene duplication is thought to be one of the major driving forces in evolutionary innovation since it introduces the ability to have novelty without the complete loss of the original function by retaining unaltered copies of proteins. Given that genes can have more than one function, sub-functionalization refers to the division of an ancestral function of a gene between the original and a duplicate. This tends to generate a division of a more traditional function of the protein with complementary degenerative mutations. While most genes are lost after a duplication event, tissue specific expression may lead to increased retention of multiple isoforms since paralogs can divide ancestral functions (sub-functionalization) or find new functions (neo-functionalization) (117, 122). If gene duplicates are redundant in function, one member of the pair will typically be silenced by degenerative mutations (113). However, if neutral mutations result in a division of labour between duplicates such that each retains a subset of the ancestral function, sub-functionalization is likely to occur. Gene function can also be context-dependent. This means that sub-functionalized paralogs may be expressed preferentially on a temporal or developmental scale, or in response to variable environmental factors. Gene expression can be divided between paralogs by reciprocal loss of cis-regulatory elements (45). Modification of ancestral function between two paralogous genes can be achieved by inactivation of functional domains or via changes within the coding sequence.

Regulatory sub-functionalization is distinguished by variation in gene expression profiles. The breadth of expression in different tissues results in variation of evolutionary rates, proteins that operate in a range of molecular environments are subjected to more evolutionary constraints. Differential expression profiles between paralogs are regulated
by transcription factor binding to distinct elements in regulatory regions of a gene. The DDC model suggests that many duplicates are conserved via mutations in regulatory regions first. These degenerative mutations are neutral, since one gene still performs the ancestral function that was lost by others. This way, functions of the gene can be subdivided between duplicates since together they can continue to perform the overall function of the ancestral pre-duplication gene (101). Post-translational regulatory changes can also contribute to functional divergence such as variation in phosphorylation between paralogs and differences in localization of the protein products.

Measuring sub-functionalization as structural functional divergence on the protein level can be more difficult. With sub-functionalization, asymmetric evolution may not be apparent at the whole protein level. The two gene copies may diverge through different fast evolving functional domains or relaxed purifying selection on a few residues. These cannot always be distinguished based on evolutionary rate alone. Because of the modular nature of proteins, analysis of asymmetric evolution and functional divergence of gene duplicates is advantageous. It can detect asymmetrical evolution that is limited to a small domain that would normally go undetected due to a large signal-to-noise ratio. If different domains are responsible for sub-functionalization, they may remain undetected at the whole protein level as opposing signals of asymmetry between these domains would cancel out in the analysis. Asymmetry detected in evolutionary rates are primarily localized within specific functional domains due to non-synonymous changes between duplicates that are clustered within a few domains rather than spread evenly across all domains of a protein (96).

1.3.4. Intermolecular coevolution

Natural selection favours evolutionary coadaptation of interacting proteins to maintain or improve physiological function (124, 140). Studies of interacting proteins have found correlated evolution of the sequences of binding partners, apparently as a result of compensating mutations to maintain specificity (47, 173). For example: many different amino acid substitutions among interfacing residues are necessary to optimize the intermolecular interaction (140). This will be present in reciprocal changes in interacting
proteins where variation in selection pressures on one will affect the other and vice versa.

These patterns in molecular coevolution are often associated with observed covariation signals. In amino acid sequences, covariation can be detected when some pairs of residues co-occur in multiple sequence alignments with greater frequency than expected. This implies that some residue pairs have higher fitness in combination than individually, so a substitution at one position will lead to a second substitution at an interacting position. Epistasis occurs when the fitness effect of one mutation depends on the state of other loci. Epistasis allows for mutations that appear neutral at one point in evolution, but permit a protein to tolerate subsequent mutations that otherwise would have been deleterious. Examples of epistasis can exist within a specific protein, or between interacting proteins. In co-evolutionary terms, selection pressures arise from functional or structural pressures acting to maintain specific subsets of residues at epistatic positions. The measurement of covariation between residues within and between proteins is commonly referred to as mutual Information.

Covariation is difficult to detect based on sequence patterns alone. Coevolution probably only occurs in a small fraction of physical interactions (112). Spatial and functional constraints other than direct interactions may be important factors in covariation signals, demonstrating a complex relationship between coevolution and selective constraints. Protein-protein interactions may impose evolutionary constraints on protein interfaces to maintain favourable interactions and avoid unfavourable interactions. The constraints on evolution preserve optimal binding affinity for biological processes. Consideration of both evolutionary history and structural conservation should be considered for effective detection of coevolution.

Evolutionary pressures may be similar only in particular interacting domains of proteins, involving only those residues involved in the interaction showing signs of evolutionary constraint to maintain function. The existence of paralogous genes adds further difficulty in analyzing coevolution. A protein with a unique interacting partner in one species may have multiple potential partners in another species. Regulation patterns should also be considered, as the gene expression level has a strong influence on the
evolutionary rate, and interacting proteins will tend to have similar levels of expression (69).

1.3.5. Current models of coevolution of protein-protein interactions

Typically, many evolutionary models have looked at complementary changes at the intramolecular level, where changes within a protein are important to proper function (149)). The importance of intermolecular interactions in protein stability and overall function has necessitated the development of multimeric protein models. Well- characterized models examining the evolution of structure-function relationships in vertebrates include members of the globin family as well as subunits of the oxidative phosphorylation pathway.

The globin family has been a classic model of evolution at the molecular level and is well characterized in terms of changes in amino acid sequence and changes in protein structure (105, 107, 129). A large amount of structural and functional data exists for members of the globin family (108). Rigorous study of these proteins has provided examples of conservation of interactions between helices that maintain the tertiary structure via common packing patterns and examples of functional divergence in sites critical to intermolecular interactions (64, 105). Tetrameric hemoglobin consists of two alpha and two beta subunits, each of which contains a heme group. For proper oxygen binding, two types of subunit interfaces are involved in a transition between quaternary structural states. Thus proper inter-subunit and intra-subunit contact regions are critical prerequisites for hemoglobin functions such as oxygen affinity and O₂ binding cooperativity (105). Myoglobin has been used to measure selection for thermodynamic stability, showing that the protein is evolving under selection pressure to maintain its folding stability (25). Within the entire globin family, the structural packing patterns that contribute to correct interactions are conserved with as little as 40% sequence identity (105). The overall structure is conserved to maintain structural stability, yet permits variation in specific interactions between various paralogs, a process that drives diversity in myoglobin function (64).
Genes that code for proteins involved in oxidative phosphorylation are encoded by two genomes: the mitochondrial genome and the nuclear genome. Cytochrome oxidase is an example of gene products from the mitochondrial genome interacting with gene products from the nuclear genome (124). In order to maintain function, coevolution between these genes that has led to intergenomic coadaptation has been examined in the literature (4). However, there is substantial evidence of differential forces that drive mitochondrial DNA (mtDNA) and nuclear DNA (nDNA) evolution (3). Residues encoded by mtDNA typically evolve at much higher rates than those coded for by nDNA (25). The evolutionary rates of mtDNA-encoded residues depend on whether they are in contact with nDNA-encoded subunits or with other mtDNA-encoded subunits (4). Interestingly, this model has revealed that these patterns in evolutionary rates cannot be explained by the thermodynamic stability of the complex. Detection of selection pressures that act on the interfaces of protein-protein complexes that facilitate functional divergence can be applied to understanding the evolution of other multimeric proteins.

Similarly to the examples provided by myoglobin and oxidative phosphorylation, the Tn complex exhibits coevolution between interacting proteins. In order to have a functional complex, three subunits are required with a large number of interacting residues between them that have roles in both stability and function. The conformational changes in the troponin complex that facilitate contraction imply that interacting residues are different at different functional conformations. While Tn is well characterized in terms of structure and function, there is little insight into the forces and constraints that shape the coevolution of the interacting subunits of this complex.

1.3.6. Evolution of the Tn family

Of the three subunits that make up the Tn complex, TnC has an independent origin from the linked TnT/TnI evolutionary history. All three proteins have been linked to specialization of vertebrate muscle types that is thought to have occurred through genome duplication events (139). All Tn subunits show tissue-specific as well as species-specific distinctions. However the muscle-specific expression pattern varies between TnC and TnI/TnT, which reflect their variation in phylogenetic history. Hence, these proteins are
typically considered separately.

**TnC**

TnC has a high degree of conservation across phylogenetically diverse groups of organisms and hundreds of millions of years of evolution (55). Because of the key role of this protein, modifications in the sequence of amino acids that change the function of TnC can alter the entire contractile reaction of the myocyte (118). TnC belongs to multi-gene family of Ca$^{2+}$ binding proteins that includes calmodulin and parvalbumin. Ca$^{2+}$ is widely used for the regulation of signaling pathways in eukaryotic cells. The EF-hand structural motif that was originally identified in parvalbumin to bind Ca$^{2+}$ (102) is found in many protein superfamilies. This largest class of Ca$^{2+}$ sensing proteins is defined by having at least one EF-hand or helix-loop-helix motif that can bind to Ca$^{2+}$. The EF-hand motif is extremely versatile in terms of function and EF-hand proteins are found in diverse conformations, domain organization and structural responses to Ca$^{2+}$ (63). While the functional properties of the Ca$^{2+}$ binding sites of these different proteins may be altered, the tertiary structure of these proteins is largely retained.

The EF hand contains a helix-loop-helix motif characterized by a sequence of twelve residues, seven of which are ligands in pentagonal bi-pyramidal metal coordination. The EF-hand motif always occurs in pairs that maintain their structural integrity through backbone hydrogen bonds between Ca$^{2+}$ binding loops. TnC and calmodulin are examples of proteins that contain four EF-hand motifs. This is conserved in two domains, each formed by a pair of EF-hands, separated by a flexible linker that have classical dumbbell structures. There is a clear difference between the conservation of the N and C-terminal parts of Ca$^{2+}$ binding loop. The N-portion is variable, and significant differences in length and amino acid sequences can occur without affecting the pentagonal bipyramidal Ca$^{2+}$ ligand geometry. This may include substitutions of the residues in Ca$^{2+}$ binding positions. The C-terminal part of the loop is constant and spacing between the residues at the $–Y$ and $–Z$ positions is conserved.

All regulatory EF-hand proteins have the ability to change conformation upon Ca$^{2+}$
binding. An example of this is the open domain conformation TnC moves to when bound by Ca\textsuperscript{2+}, which exposes a hydrophobic surface to create a target interaction site. A range of acceptable helix-helix orientations is found in EF-hand proteins and is dependent on the flexibility afforded by the Ca\textsuperscript{2+} mediated conformational change. Some EF-hand proteins become more stable upon Ca\textsuperscript{2+} binding. The function of many EF-hand proteins, such as parvalbumin in muscle physiology and calmodulin in the cytosol, depends on differences in relative affinity for calcium or magnesium. EF-hands within one protein can have different binding affinities for metal ions, their pairing results in cooperativity as a functional consequence of Ca\textsuperscript{2+} binding (52). The comparison between Ca\textsuperscript{2+} binding sites in the N-lobe relative to the C-lobe of TnC provides an example of this cooperativity.

Since TnC genes are well conserved, variants maintain overlap in basic function. The tissue-specific TnC paralogs in vertebrates provide a good example of variation while still retaining the same basic function. TnC1 is expressed in slow skeletal muscle and cardiac tissue while TnC2 is expressed in fast skeletal muscle. The affinity for Ca\textsuperscript{2+} of site II in sNTnC (TnC2) is 1.5 fold that of cNTnC (TnC1) when quantified by NMR (110). Further, in all available vertebrate sequences, TnC1 residue substitutions in site I have disrupted its ability to coordinate Ca\textsuperscript{2+} rendering it non-functional. This results in N-terminal TnC1 only operating with one functional site whereas TnC2 has retained two. In sTnC, the hydrophobic patch that opens following Ca\textsuperscript{2+} activation is much larger than the opening that occurs in cTnC (50, 148). These differences influence the strength at which the Ca\textsuperscript{2+} signal is transferred through the thin filament. Along with facilitating higher Ca\textsuperscript{2+} sensitivity, TnC2 in mammalian species does not show the decrease in Ca\textsuperscript{2+} sensitivity with decreased temperature that is found in TnC1 (74).

Vertebrate TnC2 orthologs are more variable than orthologs of TnC1 (55). This may indicate a greater scope for physiological capacities in fast skeletal muscle or reflect that proteins expressed in multiple tissues have more evolutionary constraints than those with more limited expression patterns (138). While site II, in particular, is completely conserved, the sequence of the non-functional site I in TnC1 orthologs displays significant variation. The agnathan lamprey (L. japonica) TnC1 and TnC2 share 83% and 70% identity respectively with human orthologs. The appearance of muscle-specific paralogs in the lamprey suggests TnC has played a crucial role in differentiating the functional
properties of skeletal and cardiac muscle since early in the vertebrate lineage. The duplication that established these two genes must have occurred sometime after the divergence of Urochordata (e.g., sea squirt) but before the divergence of Agnatha, sometime between 500 and 600 million years ago (MYA) (139).

Human and trout cardiac TnC orthologs show 92% identity (55). Greater variability exists between endothermic and ectothermic orthologs compared to those between endotherms. The difference in species-specific TnC1 orthologs regulates the Ca\(^{2+}\) sensitivity of force generation. The Ca\(^{2+}\) sensitivity of the teleost cardiac contractile element is significantly higher than that of mammalian when compared at the same temperature (21). Site II of TnC1 in trout has twice the Ca\(^{2+}\) sensitivity (56) and does not assume the open conformation to the same extent as that of human TnC1 (14, 53). This variability in sensitivity has been attributed to residues Asn2, Ile28, Gln29, and Asp30 (NIQD)(54). These residues are not present in human TnC1 and are located outside of Ca\(^{2+}\) binding site II. TnC1s from ectothermic species, including *Xenopus laevis*, contain at least two of the four residues identified as being responsible for the high Ca\(^{2+}\) sensitivity of trout TnC1. At their respective physiological temperatures, the structures are still very similar between trout and human. It is possible these sequence variations could counteract the effect of temperature to allow for a similar functional conformation (55).

Mammals have only two TnC paralogs, TnI1 overlaps muscle-specificity between slow-skeletal and cardiac. In zebrafish embryos two separate TnC paralogs have been identified that are comparable to mammalian cTnC: one primarily expressed in the heart (TnC1a) and the other in slow skeletal muscle (TnC1b) (151). TnC1b overexpression was able to rescue TnI1a morphilino knockdown in zebrafish hearts (151) suggesting their role may be similar, but the localization pattern suggests differences in function between the two paralogs. Yang *et al.* (2000) found no difference in TnC1a expression between cold- and warm-acclimated trout, which implies that there is only one TnC1 paralog used to trigger myocyte contraction over the teleost’s thermal range (175). However, this study focused on TnC1a and TnC2 and neglected the TnC1a/TnC1b ratio. If two TnC-like paralogs are expressed in the teleost heart, examining their expression with respect to temperature may reveal functional differences. This can also lend insight into the evolutionary development of having a pleiotropic TnC in mammals that interacts with
exclusively muscle-specific TnI/TnT.

**TnI/TnT**

While TnC has only two muscle-specific paralogs across all vertebrates, there are three for each of TnI and TnT. Expression of the three TnT genes in adult cardiac and skeletal muscles is controlled in a muscle fiber type-specific manner. Muscle-specific paralogs have diverged more than 330 MYA based on their appearance in the hagfish genome (167). Production of these tissue-specific paralogs probably occurred after the divergence of urochordates 500 MYA but before the divergence of the fish lineage from other vertebrates 440 MYA (24, 75, 76, 143). A typical pattern of fast skeletal, to cardiac to slow skeletal phylogenetic relationship is seen in muscle proteins (139). TnT and TnI isoform genes are closely linked in three tandem pairs in the vertebrate genome: TnI2-TnT3, TnI3-TnT1, TnI1-TnT2. All these genes were duplicated from one common ancestral gene (18, 91). Epitope analysis supports that TnI and TnT arose from a TnI-like ancestor protein that diverged prior to the emergence of vertebrates. The evolution of TnT/TnI genes is speculated to be driven primarily by adaptations to differentiated functions of muscle type (167). In fish, the spatial separation of fast and slow skeletal muscle brings additional interest to this specialization.

In higher vertebrates, the topology of TnI phylogeny follows a tissue-specific clade pattern in which TnI3 (cTnI) is monophyletic with TnI1 (ssTnI), which both diverge from TnI2 (fsTnI) (18). Fish and lower vertebrates do not have a specific TnI3 gene, but instead TnI genes expressed in cardiac muscle group with mammalian TnI1 (75, 143). Typically, TnI1 is transiently expressed in the heart during development of mammals (87) and TnI3 is exclusively expressed in adult hearts. However, in *Xenopus laevis* TnI1 is not expressed in the developing heart (166) which suggests that there are differential functions for TnI paralogs across vertebrates. Zebrafish express at least one TnI1 paralog in cardiac muscle (48) while trout express up to four (2). This raises the possibility of teleost TnI being pleiotropic in terms of muscle-specificity.

A distinguishing feature of mammalian TnI3 is the NH$_2$-terminal extension. This is
a hydrophobic region rich in proline residues that is critical for positioning the regulatory domain of TnC (88). This region is also an important target for muscle-type specific post-translation modification via PKA phosphorylation. Only TnI3 has the PKA target residues S23/S24. In amphibians, the NH$_2$ extension of TnI3 is not the same as in mammals, presenting as a more glycine-rich region in *Xenopus laevis* (166). Fish TnI1 does not contain an NH$_2$-terminal extension, which suggests that fish TnI paralogs are similar to mammalian TnI1 in function. This presents greater difficulty in the identification of potential cardiac-specific teleost TnI paralogs without measuring their expression levels.

Specific sequence motifs in TnI are highly conserved, including the inhibitory region and switch region (143). These regions are also critical in defining the difference between tissue-specific paralogs. For example, a major distinguishing feature between mammalian TnI1 and TnI3 occurs in the switch region (126). The presence of a histidine residue at site 132 in mammalian TnI1 has been postulated to protect contractile function during acidosis in skeletal muscle (27). In mammalian cardiac muscle, the equivalent residue at position 164 is an alanine and is involved in the loss of Ca$^{2+}$ sensitivity with decreased pH (127, 134, 168, 171). Further work showed that four residues specific to mammalian TnI3 (H164, V166, N173) are responsible for the loss of Ca$^{2+}$ sensitivity relative to TnI1 (160, 171). These residues are thought to have been an evolutionary specialization of TnI in mammalian cardiac muscle (126), however this generalization does not account for the multiple TnI1 paralogs that are expressed in fish cardiac muscle (2).

Variability in the NH$_2$-terminal region also defines TnT paralogs. The large differences in the size of TnT paralogs are due, almost exclusively, to their variable length in this region. In teleosts, the NH$_2$-terminal region is also highly diverged from that of mammals (167). Some fish species have paralogs that do not have this NH$_2$-terminal region present. This is suspected to provide the structural variation necessary for functional demands of different muscle types. The middle and C-terminal regions are conserved between species as well as between muscle-type specific paralogs. The presence of fewer exons in fish genes could suggest that the duplication of muscle type specificity occurred prior to alternative splicing capabilities. However, higher vertebrates have near identical exon-intron organization in the conserved middle and C-terminal regions (91), which suggest that the more likely scenario is intron loss in fish TnT genes.
Developmental regulation of cardiac TnT is conserved in zebrafish. Alternative splicing of NH₂-terminal exons switches from producing a high molecular weight, acidic variation of TnT to a lower molecular weight, less acidic variant as all vertebrates develop (86). There is also intriguing variability in muscle specificity of TnT across vertebrates. For example, in the toad heart ssTnT is exclusively expressed while all the other Tn subunits are the expected cardiac-specific forms (43). This results in toad hearts with faster contractile and relaxation velocities and significantly higher tolerance to afterload. Uniform TnT paralog expression in mammals is considered to be critical for rhythmical ventricular pumping but multiple variants in skeletal muscle contribute to graded twitch response (167).

1.3.7. Subfunctionalization in Tn family

TnC is a very conserved protein with very few substitutions across vertebrate phylogeny, but the presence of three paralogs in fish may have important functional implications. The function of TnC in fish cardiomyocytes may be affected by extra paralogs either directly through fish-specific variants of TnC or indirectly through variants of interacting members of the Tn complex. Molecular coevolution suggests that a change in one locus affects the selection pressure at another locus (6). Significant evolutionary constraints exist on accepted replacements in TnC due to necessary structural stability and conservation of functional sites, which results in a slower evolutionary rate. Even a minor variation in protein structure is an important consideration for Tn function (17). While TnC may be highly conserved, changes between paralogs may result in intramolecular selection pressure, which leads to differences in Ca²⁺ sensitivity that allow the fish to cope with changes in temperature. There may also be intermolecular selection pressure, where the evolution of TnI is constrained by its functional interaction with TnC.

TnC paralog expression would be expected to vary with the paralogs of the other subunits of the Tn complex it interacts with, such as TnI. In trout, TnI has seven known isoforms that may be expressed in a temperature-dependent manner (22). Tissue-specific paralogs of TnI in mammals have the ability to influence TnC Ca²⁺ sensitivity (111). The balance of a particular TnC paralog and TnI paralog may be important in the temperature variation response of the teleost cardiac contractile element. The combination of paralogs
used in the Tn complex is important for function, as seen \textit{in vitro} as cTnT and ssTnI together create less relaxation of cardiac fibers through reduced inhibition of force development (60). However, little is known about the impact of expression of multiple paralogs in teleost tissue on uniform heart contraction.

In the current literature, the presence of this additional TnC paralog in species beyond zebrafish is unknown. The effect of temperature stress can elucidate the consequences of sequence variation both to the TnC and to the Tn complex. Overlapping expression patterns of tissue-specific Tn paralogs suggests that there is a modulatory effect of interacting partners on muscle-specific function. Typically, TnC evolution is examined separately from other Tn subunits, but must have some overlapping selection pressures in order to maintain function. The troponin complex is well characterized in terms of structure and function, but is unexplored as a model of co-evolution between specific interacting domains. This dissertation will approach the functional divergence of teleost Tn as a model of coevolution.
1.4. Objectives

In this dissertation, the sub-functionalization of Tn paralogs is explored to provide insight into how teleosts achieve physiological versatility in contractile properties. Teleost Tn paralog usage with temperature fluctuations will be used to contextualize variations in cardiac function found across phylogeny and relate them to Tn protein evolution. We hypothesize that the variation of Tn paralogs will result in expression-dependent functional differences evident in correlated evolutionary patterns.

This dissertation will have three aims:

1. Determine the nature of the variation in cardiac function in teleosts.

   Chapter 3: review of the literature regarding how the teleost heart is specialized to allow for variation in contraction

   Chapter 4: exploration of the functional changes in the zebrafish whole heart due to temperature plasticity and remodeling

2. Determine the evolutionary history and context-specific sub-functionalization of the cardiac-specific TnC genes.

   Chapter 5: clarification of the TnC evolution in teleosts and regulatory sub-functionalization of teleost-specific TnC1a and TnC1b

   Chapter 6: examination of the functional characterization of structural sub-functionalization in teleost-specific TnC1a and TnC1b

3. Determine whether correlated evolutionary patterns in the interacting members of the Tn complex TnC and TnI are found

   Chapter 7: exploration of whether the functional divergence of paralogs that form a multi-protein complex in which the co-evolution of interacting partners is reflected in variation in contractile function

The co-evolution of troponin subunits is the basis for understanding the origins of variation in contractile function. This work introduces the troponin complex as a model of co-evolution of interacting proteins that are well-characterized in terms of structure and function.
1.5. References


39. Fabiato A. Calcium-induced release of Ca$^{2+}$ from the cardiac sarcoplasmic reticulum. *Am J Physiol* 245: C1-14, 1983.


Chapter 2.
Using functional divergence to characterize Tn paralogs

Gene duplication events produce paralogous proteins that can develop separate roles through sub-functionalization. While sub-functionalization may occur on a cis-regulatory level, functional divergence typically produces amino acid changes that lead to diversity in function (20, 22). This can be computed based on sequence analysis with knowledge of the role individual residues will have on overall protein function. The availability of sequenced genomes has made this comparative analysis of protein-coding genes common. Since functional constraints play an important role in the evolution of proteins, understanding the patterns of molecular evolution in protein-coding genes provides information regarding biophysical constraints. Knowledge of structure contributes to more accurate evolutionary analysis, and likewise sequence-based evolutionary information can predict the biophysical properties of proteins. A phylogenetic approach allows for predictions of functional characteristics, identifies residues of interest and provides patterns of paralog usage to help explain the co-evolution of subunits. This dissertation will integrate traditional forms of evolutionary analysis with structural information to expand upon the knowledge of Tn function across phylogeny. The focus will be on the current state of techniques used to characterize functional divergence.

2.1. Estimations of phylogenetic history

Phylogenetic trees estimate the relationships among a given set of sequences and their hypothetical common ancestors (26). Trees consist of a series of nodes connected by branches. Each branch represents persistence of a genetic lineage over time and each node represents the theoretical birth of a new lineage (such as a new paralog). Trees are inferred from sequence data, so advances in sequencing techniques and increased coverage of the genomes of multiple species allow determination of these relationships to be far more informative. The phylogenetic inference is used to estimate tree topology as
well as predict branch lengths that inform degree of relatedness. Any estimation method needs to be consistent in reporting true parameter values, efficient in regards to variance, robust even in the face of incorrect assumptions and have reasonable computational speed. Proper use of phylogenetic analysis can clarify both orthology and paralogy in gene families, but proper knowledge of statistical models used is crucial to avoid inaccurate tree construction.

Phylogenetic methods can either be distance-based or character-based. Distance-based methods, such as neighbour-joining (NJ), calculate the distance between each pair of sequences and construct the tree from the resulting distance matrix. These methods are computationally fast but can perform poorly due to large sampling errors involved with very divergent species. Typically these measures do not provide an accurate estimate of evolutionary distance when depicted as observed distances between sequences. NJ trees are best used as a starting point for more computationally-intensive phylogenetic searches but are not statistically strong enough for larger datasets.

Character-based methods, such as parsimony, and maximum-likelihood compare all sequences in the alignment for one character at a time to calculate the score for each tree (26). These are also referred to as tree-searching methods, with many trees estimated and some criterion determining the best fit. The tree with the best score is identified by a heuristic tree search that allows analysis of large data sets. Maximum likelihood (ML) and its subtype, Bayesian inference (BI) methods make use of an explicit substitution model to consider conditional probability whereas with parsimony, the model is implicit. Developments in reducing systematic errors and increasing robustness to model violations have been more important than random sampling errors in large datasets (35) and are thus important considerations in comparing these methods.

Parsimony scores a tree by the minimum number of mutations that could produce the data. The character length is the minimum number of changes required for that site. This greatly simplifies the reconstruction of the tree, since there are often many plausible pathways to explain the data. The lack of explicit assumptions makes it difficult to incorporate any knowledge of sequence evolution into tree development. Parsimony makes the assumption that the number of changes is equal on all branches of a tree, so
a mutation is weighted the same regardless of where it falls. This property makes parsimony susceptible to long-branch attraction, where two long branches are grouped together regardless of their true relationship. The simplistic reconstruction of parsimony is computationally efficient but does tend to be the least statistically consistent of the character-based methods and the most susceptible to convergence on the incorrect tree.

Maximum likelihood (ML) is based on frequency probability, which determines the likelihood something will happen based on how often it occurred previously. ML analysis used in phylogenetic inference gives the probability of the data (such as a set of aligned sequences) being observed under a given phylogenetic tree and a specified model of evolution.

\[
\text{likelihood} = \text{probability} \left( \text{data} | \text{tree topology, model parameters} \right)
\]

Hence ML analyses generate a tree that, under some models of evolution, maximizes the likelihood of observing the data (49). This produces the tree that has the highest probability of producing the observed sequences, or the probability of the data set given a phylogenetic tree using a specific model (12). Models of sequence evolution describe relative probability of various events, such as transitions versus transversions. Most models assume independent evolution of sites in the sequence so that the likelihood is a product of the probabilities for different sites. All possible mutational pathways compatible with the data are considered and the likelihood function is a consistent and powerful statistical inference. General models are relatively robust despite simplifying the processes underlying sequence evolution. An original tree is generated by applying NJ and the related BIONJ algorithms (14) to a matrix of pairwise distances estimated using maximum composite likelihood approach with a codon substitution frequency rate model for nucleotide sequences, or by applying an empirical amino acid substitution rate model for protein sequences (57). From there, a search space is initiated by evaluating alternative trees with branch swaps to determine topologies with the highest ML value. These trees are subjected to iterative rearrangement until no trees with greater likelihood can be found. The main issue with this phylogenetic analysis is its computational burden since often a single tree must be scored hundreds of times for datasets with large numbers of sequences. The number of possible tree topologies increases exponentially with the
number of starting sequences. Bootstraps are used to assess confidence, or determine how strongly the data supports each of the relationships depicted in a tree. Bootstrapping is a sampling method that pseudo-replicates the collection of data to estimate the reliability of the data. This provides a prediction about whether the same result would be seen if more data were collected, not whether the result is correct. Estimation of the parts of the trees that are supported is provided by higher bootstrap values. Maximum likelihood is a highly robust form of phylogenetic inference in estimating tree topology as well as measuring confidence of that estimate.

Bayesian inference (BI) is a form of maximum likelihood analysis that seeks the trees with the greatest likelihood given the data (29). Bayes’ theorem states that the posterior information is proportional to the prior information multiplied by the likelihood of the data. Probability calculated by BI thus refers more to uncertainty of the results based on the available evidence, making it a marginal probability over parameters. In phylogenetic inference, Bayes’ theorem is interpreted as the probability that the tree is correct given the data and evolutionary model.

\[
\text{likelihood} = \text{probability (tree topology, model parameters|data)}
\]

Posterior probabilities estimate the reliability of the model-based tree after learning about the data in an iterative process (28). The optimal hypothesis is the one that maximizes the posterior probability, with a result in the form of a consensus tree. Bayesian methods can allow complex models of sequence evolution to be explored while providing measures of support faster than ML bootstrapping by sampling trees according to posterior probabilities rather than searching for one ideal tree. These prior probabilities also allow incorporation of a priori information about the trees or parameters. Priors can add strength to the analysis but they can also be subjective and have unexpected influence on the posterior probabilities. These probabilities ideally would involve a summation over all trees and integrate over all possible combinations of branch lengths and substitution models and parameter values. However, doing this analytically is challenging. The application of approximation methods such as Markov chain Monte Carlo (MCMC) have allowed Bayesian inference to become popular for large data sets (24). MCMC effectively treats topology as a nuisance parameter by summing it over the trees. This allows measures of
uncertainty to happen much more quickly than would happen with ML bootstrapping. This series of steps forms a conceptual chain where the time spent in particular locations of parameter space provide estimates of posterior probability (24). The location in parameter space is both a description of the tree and a specification of all the parameters in the model of sequence evolution. Any probability is contingent on the model of evolution being adequate and prior distributions on the parameters being within reason.

While ML and BI are both forms of maximum likelihood, the two treat parameters in sequence evolution models differently. Many parameters are nuisance parameters of no direct interest but must be dealt with since they are found in likelihood equations. ML depends on joint estimation and finding the highest point in the parameter landscape, while BI uses marginal estimation or measures volume under a posterior probability surface. The use of marginal estimation in BI integrates out nuisance parameters using MCMC. Integration of parameters requires weighting them according to their posterior probability, which may be considered subjective. This differential approach to parameterization can result in different tree topologies from ML and BI: a phylogenetic tree is only as good as the alignment. In order to be confident in an alignment, it is necessary to have an average sequence identity of greater than 50% to ensure accuracy. Ideally, functional similarity should be derived from the functionally constrained columns of a multiple sequence alignment, while evolutionary distance should be inferred from both constrained and unconstrained positions (39). Aligned columns with gaps are conventionally disregarded in phylogenetic analysis due to increased bias in branch length calculations. A multiple sequence alignment cannot determine whether an unaligned gap is due to an insertion mutation in sequence or a deletion mutation in a sequence. The variation in insertions or deletions (indels) correlates with the variation present in aligned sequences. Therefore, it does not add further statistical consistency to the analysis and thus is treated as missing information in multiple sequence alignments (52).

A final consideration in phylogenetic analyses involves model testing. Character-based analyses are dependent on the assumptions made in the model used in the analysis. Modeling assumptions that poorly fit the observations lead to error, hence having the most realistic model available is important. This can involve adding estimations of parameters – such as allowing transitions and transversions at different rates. However,
adding parameters creates greater uncertainty in estimates so it is preferable to use a minimum number of parameters. The robustness of the estimates of the evolutionary parameters under different models of substitutions and assumptions about the distribution of evolutionary rates among sites are predicted by goodness-of-fit as measured by Bayesian information criterion (BIC) and corrected Akaike information criterion (AICc). The AIC provides a score of the likelihood of the model being penalized by a number of free parameters. Both BIC and AICc tend to select more complex substitution models than the true model, but it is highly probable they will provide an accurate model selection. Typically the correct model is selected with a higher frequency by BIC than AICc, so programs such as MEGA6 base model prediction on this index. The choice of criterion to best-fit models is complicated so both indexes may be necessary for accurate predictions. For nucleotide substitutions, five nested models can be used for ML, but for amino acid substitutions, six models with and without empirical frequencies are preferable. An evolutionary model providing good fit to the data is also selected based on computational expense, resulting in the simplest model being preferentially selected.

The method of phylogenetic analysis that should be used depends on the data being analyzed, Model-based methods such as ML and BI are generally more robust and consistent than the more simplistic parsimony methods. In this dissertation, both ML and BI are used to examine relationships between troponin paralogs. Both ML and BI allow rates of evolution to vary between lineages and sites. The measure of accuracy in both analyses is different. BI answers biological questions in an easier to interpret manner through posterior probabilities, which indicate degree of uncertainty of how correct a tree is. Nonparametric bootstrapping of ML trees results in a statistically more difficult explanation since it does not give an indication of whether a result is correct, but rather an idea of the repeatability of the results. However, posterior probabilities in BI are often subject to over-inflation and have greater sensitivity to model bias. As well, the incorporation of prior knowledge may incorporate bias into branch lengths resulting in too long trees. ML is much more conservative in bootstraps and includes an algorithm with a branch length optimization. ML is well-suited to distantly related sequences, while BI used with large data sets where the amount of information would be too long computationally for ML. With shorter proteins such as TnC, and looking between species rather than populations, the computational time does not exclude ML. However, unlike BI posterior
probabilities, the conservative bootstraps may not distinguish between closely related species for analyses across vertebrates. For the analysis of highly conserved proteins such as Tn, the trends described by both methods can be mutually corroborative.

### 2.2. Phylogenetic approach to understanding the fate of duplicates

Molecular phylogenetics are fundamental tools in cross-species comparison as it can inform us about which genes should be compared between species. Genes related by speciation are designated as orthologs whereas paralogs are related by gene duplication. Both are examples of genes derived from a single common ancestor, necessitating the availability of specific information about evolutionary relationships to distinguish the two. The recognition of true orthologs may be complicated by evolutionary processes such as gene gain/loss, gene transfer and protein domain emergence/rearrangement. Gene duplication serves as the source of multiple genes and is often used to infer orthology or paralogy. Since the doubling of the gene content in fish due to the teleost-specific 3R between 320-250 million years ago (6, 53), a loss of 80-85% of the duplicated genes has been estimated to have occurred (4). This makes it difficult to predict how many paralogs to look for, and to classify whether certain genes are orthologous in fish. It becomes a more complicated issue with the polyploidization that is associated with the teleost-specific 3R, as well as the asymmetrical sequence divergence of duplicates. Paralogs may be retained in some species but exist as only a single copy in others, demonstrating lineage-specific paralogs; or orthologous genes may have been retained as paralogs throughout a particular suborder.

Since orthology and paralogy are coupled to specific types of evolutionary events (speciation and duplications, respectively) classifications usually involves phylogenetic analysis. Orthologs typically perform equivalent functions in the respective organisms, based on their role in their shared last common ancestor, whereas paralogs undergo functional diversification. It follows that greater similarity is seen in sequences of orthologous genes relative to each other than they are to any other genes from the
compared genomes, leading to protein products more highly conserved than those of similarly diverged paralogs. Assigning orthology can be done on a gene-by-gene basis using a combination of phylogenetic methods and available data on conserved synteny or gene linkage. Large-scale analysis of syntenies requires chromosomal-level genome assemblies to determine conservation of gene order and composition in multiple chromosomal segments (30, 36). Chromosomal information is currently available for several teleost species such as the zebrafish, medaka and Fugu, as well as more basal vertebrates such as the spotted gar. Most current, second generation genome assemblies are fragmentary scaffold-level assemblies which often have gaps in the placement of particular genes. Within particular lineages with poor taxon sampling, inclusion of more species could allow for retrieval of the genes that later underwent loss or duplication in another particular lineage, except in cases where lineages have much lower species diversity. This can also lead to long branches in which no intermediate state is captured (37), rendering the phylogenetic analysis less powerful.

Using model fish species for a Basic Local Alignment Search Tool (BLAST) search against non-model fish is also problematic. In addition, there is no requirement that orthology is a one-to-one relationship. Lineage-specific reshuffling of teleost genomes also leads to further confusion of orthologs and paralogs. Retention of specific protein domains may be the only evidence left for orthologous genes. Chromosomal rearrangements may have occurred over a long evolutionary period after a duplication event, so traces of tandem duplications may have been lost for many genes. Since many paralogs are yet undefined in the fish genome, it can lead to comparison of non-orthologous genes when no further information is provided. Zebrafish, having the most detailed annotated teleost genome is phylogenetically isolated from most other species as a cyprinid. However, similar gene pattern localizations can predict the likelihood that either tandem gene duplication events or whole genome duplication events produced the resulting genes.

Conserved synteny in teleost fish genomes is generally accepted as providing firm evidence of whole genome duplication (WGD) in its basal lineage (1). Similarly, the pattern of gene retention in a synteny block that is conserved between species is expected to indicate orthology. Gene order within syntenic blocks may be altered by intrachromosomal rearrangement, while the size of these segments may be affected by subsequent
translocation of parts to other chromosomes (32). A high number of conserved syntenies are retained in the genomes of a range of species, interspersed with areas of rapid change where rearrangements have significantly altered the configuration and chromosome numbers (47). Conservation of synteny alone for this reason may not be enough to determine orthologs. Certain gene families, such as the globin genes, show high asymmetry in conservation of the gene repertoire between lineages (25). For this reason, localization of conserved gene structure should be combined with phylogenetic analyses as well as investigation into functional relevancy of the genes.

2.3. Quantification of variation in evolutionary rates

Measuring selective pressures allows us to quantify variations in evolutionary rates leading to sub-functionalization. Evolutionary distances are measured as expected substitutions per site per unit time, or percentage divergence. This standardized approach can lend insight to not only the degree of divergence between proteins but to the protein divergence location. In adaptive protein evolution, phylogeny can be used to infer cases of accelerated amino acid change. Phylogenetic Analysis by Maximum Likelihood (PAML) algorithms can estimate parameters of sequence evolution by comparing phylogenetic trees (58). Codon models are used to distinguish between synonymous mutations, which are likely to be neutral, and nonsynonymous substitutions, which are exposed to the action of selection. If synonymous and nonsynonymous substitutions accumulate at the same rate, neutral evolution is occurring where selection on the protein has no effect on fitness. If nonsynonymous substitutions accumulate slower than synonymous, it suggests that nonsynonymous substitutions are deleterious and thus a purifying selection is occurring. If the accumulation of nonsynonymous substitutions is faster than synonymous substitutions, this suggests positive selection where the amino acid changes are advantageous. Comparison of the nonsynonomous (dN) and synonymous (dS) rates as an omega ratio is used as an indication of the direction and strength of natural selection (34).
Codon based models account for variation in the omega ratio of dN/dS among sites (17). Modeling evolution among codons uses a maximum likelihood estimation of the relative rates of nonsynonymous and synonymous substitutions as a measure of selective pressure on the protein. These models must also take into account nuisance parameters that may affect estimates, such as biased nucleotide substitutions; for example, transitions being more common than transversions. Simple use of the omega ratio to distinguish positive selection between two sequences is not usually successful, because of the averaging effect across all sites. Much effort has been taken to develop models that can detect positive selection affecting specific lineages or individual sites. Branch models use different omega ratio parameters for different branches of the phylogeny to detect positive selection acting on particular lineages (56, 62) while site models allow the omega ratio to vary among codons to detect positive selection in particular sites (45, 60). Branch-site models detect positive selection that only affect a few sites on pre-specified lineages (59).

Algorithms within programs such as PAML rely on likelihood ratio tests (LRT) to detect adaptive evolution. Adaptive evolution is permitted in a model that is compared to a null model where adaptive evolution is not allowed. These tests are conservative as long as the convergence is carefully checked. Type I error occurs when the null hypothesis is rejected when it is true, usually seen as the false identification of positive selection. If the model allowing adaptive evolution explains the data significantly better than the null model (greater LRT), the empirical Bayes approach can be used to identify which sites are likely to evolve adaptively. The empirical Bayes approach relies on estimates of model parameters, which are susceptible to sampling errors in small data sets – therefore the Bayes empirical Bayes (BEB) approach is used with low false-positive rates (61).

2.4. Detecting functional divergence beyond evolutionary rates

A change in evolutionary rate may indicate that functional divergence has occurred. Analyses based on the quantification of residue variability in paralogs within species, as well as orthologs among species, help clarify functional amino acids. The
underlying principle of Gu’s probabilistic model is that functional divergence after gene duplication is correlated strongly with the change of evolutionary rate (20-23). On an individual residue level, the importance of conservation can be due to many different factors, including structural stability, protein-protein interaction, protein-DNA/RNA interaction, ligand binding site and maintenance of protein functions. Amino acid replacements tend to be more divergent (stabilizing or destabilizing structure) compared with amino acids evolving under a neutral process, where the function is preserved (7). True selective pressures must be tested for functional differences created by residue shifts (44). Functional divergence (FD) between duplicate genes can be used as a measure to link change in biophysical properties in the structure of a protein to evolutionary selection pressure on that particular site.

Statistical tools from the DIVERGE software package detect residue shifts responsible for divergence. Selection of residues using this computational method is prioritized based on functional constraint patterns extracted from multiple sequence alignments (23), linking phylogenetic information with amino acid shifts. The site specific profile based on posterior probability to predict critical residues between two gene clusters can help make predictions for further measurement. This is further mapped onto existing structures of homologous proteins. Functional divergence is classified into three types. Type 0 refers to amino acid patterns that are universally conserved, implying residues that are important for the common function of all member genes. Type I amino acid patterns are highly conserved in one gene but highly variable in the other gene, implying these residues have experienced altered functional constraints. This often involves different evolutionary rates between duplicate genes. Type II amino acid patterns are highly conserved in both genes but have very divergent biochemical properties, implying that these residues may be responsible for functional specification. Finally, Type U amino acid patterns in many residues are difficult to discern and remain unclassified. The ability to identify amino acid sites responsible for functional diversity from sequence analysis is cost-effective in terms of time and resources, and lends itself to predicting what should be tested further by experimentation. Comparative studies of gene families will also help elucidate the functional significance of single nucleotide polymorphisms in coding regions. This is especially important for highly conserved proteins such as those in the troponin family, where homology between structures is high.
2.5. Quantifying sub-functionalization of gene duplicates

After duplication, the retention of gene products is based on constraints set by expression level and functional interactions. Regulatory sub-functionalization is described by the control of protein expression, either transcriptionally or post-transcriptionally. Transcriptional regulation changes expression patterns and is relatively easy to measure. This regulation is intrinsically linked to function, hence evolution can drive functional divergence of proteins by changing their expression patterns, even if overall function is conserved. Many genes may have constitutively low expression levels and limited spatiotemporal patterns, so it may be difficult to determine if these are functional or not. Cis-regulation patterns drive functional divergence of genes (8). Gene expression is primarily regulated by transcription factors via cis-regulation, hence regulatory sub-functionalization can be quantified by expression patterns. With genome duplication, it is more common to see divergence in cis-regulatory regions rather than in coding sequence. The lack of conservation of non-coding sequence may make patterns difficult to follow, but the impact can be detected using expression levels. Many gene pairs resulting from WGD show evidence of changes in regulatory control and novel expression domains (33); spatially in tissue-specific patterns, temporally in developmental stage patterns, or in response to environmental perturbations such as temperature.

High-throughput quantification of gene expression has been an important component of expression profiling. While microarrays can compare a large number of genes, quantitative PCR is widely used for its accuracy in providing expression profiles for specific genes. Fluorescent-based qPCR has a large dynamic range of quantification, high accuracy, sensitivity and throughput capacity. The fluorescence emitted during the reaction is monitored as an indicator of amplicon production at each PCR cycle rather than endpoint detection as in normal PCR. The first significant increase in the amount of PCR product is correlated with the initial amount of target template. With lower the RT-PCR template amounts, the more amplification cycles are needed to reach threshold fluorescence. The relative mRNA quantification determines the amount of target mRNA in samples relative to each other. For most applications, calculation of the absolute amount of the measured mRNA is unnecessary. To compensate for differences in template input
quality and quantity, the target mRNA amount in each sample must be normalized to one or more internal controls. These internal controls are usually abundantly and constantly expressed housekeeping genes but these should be checked for consistency in every test since many of the commonly used ones are least reliable (54). In order for this relative quantification to be accurate, the efficiency of the PCR reaction should be 90-110% for both the standard and the target as determined by the slope of the log linear phase. A number of variables can affect the efficiency of the PCR, including the length of the amplicon, secondary structure and primer design. This has rendered consistency in the set-up of experiments important for both reproducibility and valid interpretation of results (5). This genomic expression data may be used to infer information about the presence of cellular proteins, but proteomic information describes the actual content, which ultimately determines the phenotype. For some species and under some conditions, there is a good correlation between the measurement of the transcriptome and the proteome; however, such correlation is by no means universal due to post-transcriptional and post-translational regulation. Monitoring of the expression of protein products via mass-spectrometry based techniques such as multiple reaction monitoring (MRM) can be used to link genomic information with proteomic expression (11).

Variation in regulation patterns is typically linked to changes in the function of the protein products. Quantifying structural sub-functionalization can be more difficult, in part because of the high conservation of structure as well as the potential overlap in basic function. This renders simple coding sequence comparison insufficient for detection of structural divergence. Quantifying divergence can be done using a theoretical/computational framework, but basic homology modeling on its own may also not reveal true variation in function (31). The Rate4Site program on the ConSurfDB server calculates evolution conservation profiles of protein structures to determine functional importance (16). This is based on the assumption that protein function is mediated through clusters of evolutionarily conserved amino acids located in close proximity to each other. This must be based on existing structure. Proteins must be sufficiently stable to fold but often need to be locally unstable to function (51). Stability-based or native state models are used to understand site-specific constraints since patterns of substitution create variability in stability with divergence (3), and variability in stability leads to changes in
protein activity (9, 10). The changes that influence stability can be accurately inferred from the degree of divergence between protein sequences (2). This can be complemented by measures of stabilization, such as indexes of hydrophobicity like the relative solvent accessibility (RSA) (43). The solvent inaccessible cores of proteins are closely packed, more hydrophobic and more conserved than surface regions (43). This is seen as the primary structural constraint on amino acid substitutions. Hydrophobic effect contributes greatly to the stability of the folded state of the protein and hence has been one measure used as a correction factor (55). Another measurable factor in protein stability include hydrogen bonds within alpha helices and beta sheets or via side chains. These bonds increase packing density and through the creation of stronger van der Waals interactions in a protein have a comparable contribution to protein stability. Accounting for stability has been shown to improve the fit of many models used for phylogenetic analysis (2, 43) though it has been difficult to pinpoint one specific property or measure (10).

Multiple models exist to mechanistically describe biophysical constraints on protein evolution. The stress model, or active state model, incorporates stability with conformational flexibility, essentially balancing RSA with mean square deviations (RMSD) (27). The stress model relies on a measure of proteins assuming the correct conformation rather than a more stable state, through packing density scores. Mutations of packed sites both stress and destabilize the active conformation of a protein (9). Single site substitutions perturb its interactions with other sites, destabilizing the active conformation and thus influencing the likelihood of that particular substitution being retained. Weighted contact number (WCN) scores based on side chain packing have been found to be proportional to the amount of destabilization (40). Hence, WCN can be used as a measure that correlates with site-specific evolutionary rates (42, 63). In this way, this model can be viewed as a bridge between biophysical parameters and evolutionary selection. There is debate whether RSA or packing density is the most important factor in constraining evolution of a particular residue (63). In addition, these summary statistics for biophysical parameters are rough estimates based on static structure and have difficulty accounting for the dynamic nature of protein structure as well as the interactions between residues within the protein. Utilizing a frustration index (frustratometer) provides an estimate of an energetic score (entropy) due to substitution, either conformational or mutational based on its effect on conformational flexibility or stability (13). Changes in energetic frustration
are linked to functional regions of proteins, or sites that can be linked to functional changes. This index utilizes homology models or acts as a summary score of dynamic simulation for a particular residue, giving a measure of the functional constraints on the evolution of protein energetic landscapes. This can provide a link between computationally expensive molecular dynamic structures with phylogenetic analyses to explain the importance of functional divergence at a particular residue.

By using a phylogenetic approach, predictions can be made on structural sub-functionalization prior to more extensive functionalization. In this manner, a more efficient approach can be taken. For example, only a subset of sequence substitutions in TnC would be expected to have an effect on Ca\textsuperscript{2+} affinity. Rather than considering all possible substitutions in all orthologs to determine the extent of functional divergence, the experiment may display increased focus by predicting the nature of the sub-functionalization for particular residues. Specific questions can be addressed using characterization techniques such as NMR spectroscopy or Ca\textsuperscript{2+} binding experiments with fluorophore-labelled constructs (15). A well-characterized structural model allows the investigation of the effects of divergence between paralogs more effectively than phylogenetics alone.

### 2.6. Current limitations in the phylogenetic approach

The approaches we have discussed are predominantly sequence-based. However, evolutionary-rate variation within sites depends on functional and biophysical properties of the protein. The simplest models for phylogenetic reconstruction are based on the assumption that protein sites evolve independently and identically. These include empirical substitution matrices such as JTT and WAG models used for both maximum likelihood and Bayesian tree building. This assumption of independence among sites ameliorates many complications in computing likelihood, but ignores the information contained in the protein structure. Protein folding and function cause interdependence between sites in protein sequences. In fact, coevolving residue pairs can provide information on the functional conformational landscape of protein structures (31). The ratio
of nonsynonymous to synonymous substitutions (dN/dS) have been shown to vary depending on the position of each residue in a 3D structure, with surface positions evolving faster than buried and highly constrained core residues (18).

The current methods of phylogenetic inference rely on substitution matrices that can roughly consider some amino acid properties, such as the tendency for conservative substitutions where polar amino acids to be substituted by other polar amino acids rather than hydrophobic. Many substitution models consider the basic physicochemical properties of amino acids which constitutes a large amount of data to fit needed parameters (38). However, they do not account for the local structural environment of a given residue position, such as backbone conformation and hydrogen bonding patterns, which might constrain evolution (55). These are often used in simulations but less established for phylogenetic inference because of computational burden. Good predictors of residue-residue physical interactions in the tertiary structure of proteins tend to be indirect correlations related to phylogeny in multiple sequence alignments (2). Statistical analysis of genomic data has, in turn, predicted information about protein folding and function (31, 41, 50). Traditional quantification of evolutionary divergence, including PAML measures of dN/dS, assume all sites evolve independently, making it important to further consider functional divergence with respect to structural constraints (48). Sites that contribute to functional divergence have a greater effect on protein structure than sites where functional divergence is not detected (19). For this reason, integrating structural constraints into phylogenetic estimations has been considered a logical next step in more accurate reconstructions of evolutionary history.

Several methods have been used in the past to better identify sites, integrating the use of phylogenetic and structural information (9, 43, 46). Simple models of protein folding allow protein evolution simulation but are difficult to implement with standard programs for phylogenetic inference (2). Current integrative techniques have focused on linking phylogenetic analysis with information on protein stability. Accounting for stability has been shown to improve the fit of many models used for phylogenetic analysis (2, 43). The traditional model of coding-sequence evolution can be combined with the stability information obtained from the 3D structure of the evolving protein, specifically the relative solvent accessibility (RSA) (43). The stability model takes into account the ability to fold,
whereas the stress model takes into account the probability of adopting the correct conformation (10). Both protein stability and active conformation constrain site-specific sequence evolution in proteins (9). It is unclear whether RSA or packing density is the most important factor in constraining evolution of a particular residue (63).

The use of simplified models allows protein structures in the PDB database to be combined with the computation likelihood of substitution model run in PAML. Stability of the native state is an important determinant of protein evolution, but must be combined with the likelihood of being in the correct native state to ultimately examine selection on interactions between sites. Models discussed above do not account for interactions between residues, but residues in structures do not evolve independently. Protein-protein interactions are not always constant, especially with the necessity of transience in certain interactions for function. In addition, these summary statistics for biophysical parameters are rough estimates based on static structure and have difficulty accounting for the dynamic nature of protein structure as well as the interactions between residues within the protein (10). In the future, evolutionary models should incorporate dynamic models of protein structure utilizing molecular dynamic simulations (MD) for higher accuracy of prediction. However, computational errors incurred with too many parameters involved in the evolutionary models ultimately weaken evolutionary analysis. Hence, in this dissertation, we focus on a combination of scores from both phylogenetic tests (PAML, FD) with dynamic structural models (MD) rather than utilizing a single summarizing model for both.

The fields of structure based models and statistical analyses of genomic information have often progressed independently. The integration of these approaches can provide far more information on the overall function of the protein. Current phylogenetic methods do not account for intermolecular forces, and many known structural models may not be able to account for all interactions. In Tn, knowledge of how interactions shape function makes the understanding of the coevolution of interacting subunits crucial for understanding sub-functionalization. Sequences of orthologous proteins that form stable complexes diverge more slowly than those proteins not known to be involved in interactions (10). This dissertation relies on the utilization of phylogenetic comparative methods in combination with dynamic structural models to apply evolutionary
theory to knowledge of protein evolution. Combining the use of phylogenetic inference with knowledge of structural interactions provides mutually corroborative information that strengthens our understanding of the co-evolution of interacting domains. The use of a well-characterized structural model allows us to investigate the effects of divergence between paralogs more effectively than phylogenetics alone, and the use of evolutionary selective pressures allows prediction of important residues that produce variation in the interaction.
2.7. References


52. Truszkowski J, and Goldman N. Maximum Likelihood Phylogenetic Inference is Consistent on Multiple Sequence Alignments, with or without Gaps. *Syst Biol* 2015.


Chapter 3.
Variation in cardiac contractility in fish

This chapter was modified from the following book chapter:

In order to better understand the variation that exists in cardiac contractility across vertebrates, consideration needs to be taken for the specific anatomy and functioning of the hearts of more basal vertebrates. The teleostian zebrafish is a widely used non-mammalian vertebrate model in developmental genetics and functional genomics as well as in cardiac studies (4). There is strong similarity between teleost and human cardiac action potential morphology. Basic excitation-contraction (E-C) coupling is also conserved in fundamental features such as Ca\textsuperscript{2+} sparks and the ability to trigger Ca\textsuperscript{2+}-induced after-depolarizations. However despite these similarities, the zebrafish is ectothermic with a morphologically distinct heart. Fish encompass a broad range of species with great diversity in anatomy and physiology. Because of the broad environmental conditions ectothermic fish are exposed to, critical processes such as cardiac E-C coupling must be adaptable to allow for the necessary cardiac scope. The contractile properties of the two main chambers of the fish heart, the atrium and the ventricle are likely to be important in achieving physiological versatility. The evolution of variability in cardiac contractility is an important consideration in ectotherms. This chapter is adapted from a review (32) that synthesizes research on the differing roles of the atrium and the ventricle across teleost species.

### 3.1. The fish heart

Across vertebrates, the fish heart is relatively simple in terms of structure. The heart of teleosts is unique in structure, composed of a series of four chambers: sinus venosus, atrium, ventricle and bulbus arteriosus. The two chambers that act as pumps are the atrium and ventricle, a simplified version of that is seen in tetrapods. These chambers develop from a simple linear tube (76) and differ not only morphologically but also physiologically with diverse characteristic rates of contractility (90). The design of fish hearts is reflective of the needs for oxygen delivery to working skeletal muscle in an often oxygen poor environment. The heart is the first definitive organ to develop embryologically and become functional, as any later survival depends on its proper function. The vertebrate-specific development of multiple chambers with enough muscle to generate higher systemic pressures allows for perfusion of larger and more complex tissues. The
simplified teleost heart allows for more flexibility on both the morphological and functional level, reflecting the diverse environmental challenges that ectothermic teleosts face.

3.1.1. General morphology

In fish, the heart is arranged in series with the ventricle primarily filled by contraction of the atrium (19), rather than the mammalian system in which the thin-walled atrium contributes only a small amount to ventricular filling under resting conditions. While the atrium and ventricle compose the main muscular components of the fish heart, there are other chambers with important roles in guiding appropriate blood flow through the heart. Generally in fish, blood flows through the sinus venosus to the atrium, then to the ventricle and out to the bulbus arteriosus (Figure 3-1). More recently, the atrioventricular region and conus arteriosus at the anterior end of the ventricle have been identified as morphologically discrete segments present throughout modern teleosts as well (53, 91).

Figure 3-1  Simplified morphology of a typical teleost heart.
SV refers to the sinus venosus while BA refers to the bulbus arteriosus. CA refers to the conus arteriosus region. The red arrows denote the direction of blood flow.
As with most vertebrates, the cardiac tube is S-shaped with a dorsally positioned atrium and ventrally positioned ventricle. This allows for appropriate momentum of the bloodstream towards the arterial pole. In many respects the fish heart represents a good developmental model, showing the primitive state of the more complex heart arrangement in higher vertebrates such as mammals.

Blood first enters the thin-walled sinus venosus, an independent non-muscular chamber in fish that acts as a drainage pool for the venous system (31). In humans, this chamber only exists during embryonic development (122) but in fish the continued presence of this chamber is suspected to have allowed the development of the atrium as the principal driving force for ventricular filling (57). The composition of the sinus venosus varies greatly between species, from primarily connective tissue in *Danio rerio* to mainly myocardium in *Anguilla anguilla* to even smooth muscle cells in *Cyprinus carpio* (25, 89). Blood then moves into the atrium, an active contractile chamber in fish that again shows considerable variability in size and shape between species (25). One general trend seen throughout teleost species is that the atrium is similar in volume to the ventricle, whereas the atria are distinctly smaller in mammals. The volume similarity in the two main muscular chambers in teleosts is indicative of the importance of the role of the atrium in filling the ventricle. The atrium as a chamber is still less developed muscucally than the ventricle and is possibly representative of a more primitive heart (51, 108). Separating the atrium and the ventricle is the atrioventricular region, which is a ring of cardiac tissue supporting the AV valves. The degree of isolation of this region from the surrounding musculature varies between species (54) but in any form it plays a critical role in regulating the pattern of electrical conduction. While comparable in overall volume, the ventricle has a mass and a wall thickness up to five times that of the atrium (25). Diversity of structural organization, histology, coronary distribution and relative mass is found between species. This variation reflects the fact that different modes of heart performance are attributed to different ventricle types since this chamber is responsible for the active cardiac output (25). From the ventricle, blood moves into the outflow tract leading into the beginning of the dorsal aorta. In more basal fish there are two segments: the proximal muscular conus arteriosus and the distal arterial-like bulbus arteriosus (55). While later teleosts were originally thought to have lost the conus arteriosus, more evidence supports its presence as a distinct segment in many species (53, 91). The bulbus arteriosus is more prominent and
serves as the main component of the outflow tract. This elastic chamber expands to store large portions of the stroke volume, and gradually releases a steady non pulsatile, non-damaging flow to the gill vasculature (25). The role of these segments is variable between species.

The AV-region, bulbus arteriosus and conus arteriosus are unique to fish though homologous regions may be seen in early embryological stages of higher vertebrates. The atrium and ventricle are present across vertebrates, though their roles in fish may be unique and thus their myocardial structure is distinct. The teleost myocardium is composed of two distinct layers: compact and spongy associated with vascularized and lucunary blood supply respectively (1). The compact layer is comprised of typically well-organized fibers composing well-arranged bundles covering the inner ventricle. The spongy layer is made up of long finger-like branching cells in close proximity to luminal fluids. The atrium tends to be primarily spongy, with more open space between fibers and greater variability in orientation than the ventricle (89). The external rim is myocardium surrounding a complex network of pectinate muscles. Collagen provides support, being predominant both in surrounding the myocardium and the internal trabecular network (53). Concurrently, while the ventricle composition varies greatly between species, the periphery of the ventricle tends to be primarily compact tissue with the luminal being primarily spongy (56).

Not all fish species have distinct layers in the ventricle to the same degree as the salmonids and scombroids in which most of the characterization of the chambers has been done. For example, the relatively sedentary flatfish has a heart with virtually no lumen, composed almost entirely of spongy tissue (72). These ventricular phenotypes have been subdivided into subclasses based on histology and blood supply (Type I-IV)(1). A Type I ventricle consists of an entirely spongy ventricle, while Type II has both compact and spongy myocardium but with coronary vessels restricted to the compact layer. Type III and Type IV have vascularized trabeculae, with coronary vessels penetrating spongy and compact myocardium. Types III and IV are distinguished from one another by Type III having less than 30% of heart mass composed of compact myocardium (exemplified by many elasmobranchs) and Type IV representing ventricles with greater amounts of compact myocardium (1). The composition of the myocardium will have an important
impact on the ability to modulate the cardiac performance. The zebrafish ventricle has a distinct compact outer layer and inner spongy layer with numerous trabeculae and is representative of a Type IV phenotype (73).

3.1.2. Cardiomyocytes

At the cellular level, variations exist between the atrial and the ventricular myocardium in teleosts. In general, cardiomyocytes are more cuboidal in the ventricle compared to the more squamous shape seen in the atrium across teleost species (82). Ventricular myocytes from ectothermic vertebrates more closely resemble myocytes from neonatal mammals rather than adults in terms of structure and function. They are long (>100 μm), thin (<10 μm) and generally lacking in t-tubules (trout- (25, 112); neonatal rabbit - (52)). These myocytes appear under electron microscopy to have less sarcoplasmic reticulum (SR) than mammals (88) but some peripheral couplings with sarcolemmal (SL) membrane are observed (117). Atrial muscle cells contain well-developed peripheral couplings but no t-tubules in teleosts (84). This is also true in mammals to a certain degree (75). The narrow extended shape of teleost cardiomyocytes from both chambers decreases the diffusion distance from the sarcolemma to the center of the cell (22, 113). This high surface-to-volume ratio of the cells may be related to increased efficacy of the E-C coupling without the presence of t-tubules being necessary for Ca²⁺ delivery from extracellular sources.

3.1.3. Lifestyle demands influence morphology

The morphological differences between species are reflective of the diversity of lifestyle demands for different teleosts. Active fish have a larger relative heart mass than that of less active species. As an example, the relative heart mass in the skipjack tuna is 10 times larger than in the flounder (12). However, a similar atrium-to-ventricle mass ratio exists in species with very different lifestyles (0.2 in active Atlantic salmon, Atlantic cod and inactive winter flounder - (72)). More active fish also show a difference in whole heart shape with active fish (e.g. salmonid and scombrid families) showing pyramid-shaped
hearts whereas inactive fish hearts are more rounded and sac-like in shape (87). This shape refers primarily to the ventricles, where a pyramidal shape normally correlates with a thicker myocardial wall and high cardiac output. This allows for high levels of stroke work and the maintenance of relatively high blood pressures (25).

Generally, a larger amount of compact myocardium is found in fish with greater cardiac output demands. With pyramidal ventricles, the myocardial wall is normally formed primarily by the compact layer, which is complex and thicker in these higher performance fish (25). However, this phenotype is also found in the relatively sedentary Antarctic teleosts (87). This suggests that shape and degree of compact myocardium do not always correlate with lifestyle and in fact the shape of the chamber does not always correlate with the inner architecture (102). This active phenotype may also display plasticity, as inactive trout will display an abnormal rounded shape (15, 30). The proportion of compact myocardium has also been shown to vary with seasons (25) and overall growth of the fish (13).

More compact tissue in the ventricle generally requires more perfusion of oxygen, meaning active fish require more developed coronary circulation for increased cardiac output. In general, sluggish species lack this coronary circulation and can rely completely on oxygen, contained in venous circulation (103). Hence thicker ventricles seen in more active fish have a more developed coronary distribution to the compact myocardium, and a greater percentage of compact myocardium. Very active fish like scombroids and elasmobranchs also have a coronary supply to the spongy myocardium and atrium (21). However, this increased demand for oxygen supply for cardiac function means that these species are at a disadvantage in hypoxic water. The more primitive hagfish heart composed almost entirely of spongy myocardium can meet cardiac ATP requirements for resting conditions even in anoxic water (27).
3.2. Filling patterns

Atrial filling occupies a substantial proportion (up to 75%) of the cardiac cycle resulting in a volume comparable to the ventricle. Substantial debate exists in the literature as to the mechanism of both atrial and ventricular filling, with discrepancy in results due to experimental set-up and species variability. Though non-invasive echocardiography has greatly contributed to research on flow patterns, it requires anaesthetization and restraint of the fish. Because the heart is acutely sensitive to venous pressure, and venous pressure is acutely sensitive to gravitational forces and extramural pressure, considerable variation is seen from study to study. Doppler echocardiography has been used to study intracardiac flows, but its use in fish has predominantly been limited to surgically implanted probes invoking the possibility of error-introducing calibration after removal from the animal (107), and is only possible in larger fish. Conventionally, cardiac hemodynamics have been measured by invasive catheterization techniques in fish (21, 26, 41). However, these involve manipulation in the in situ environment to mimic in vivo conditions leading to inconsistencies in the background. Certain studies have shown a more monophasic filling of both atrium and ventricle via a vis-a-tergo mechanism (e.g. trout - (74)). In this model, ventricular pressure is derived from arterial flow and venous capacitance is the primary determinant of stroke volume (19). Other studies have demonstrated a more biphasic filling pattern via a vis-a-front mechanism in which cardiac suction plays a prominent role in ventricular pressure (e.g. elasmobranchs - (65)). Even triphasic patterns have been observed, with a third wave of filling possibly involving sinus contraction (e.g. leopard shark - (65)).

These filling patterns of the chambers may be both a reflection of phylogeny and external conditions. Less active fish such as the sea raven require above-ambient filling pressures that may require vis-a-tergo filling (26). The venous pressure control mechanisms are not seen to be as developed in elasmobranchs as in teleosts resulting in a greater necessity of vis-a-front filling (86). Meanwhile with increased stroke volumes during exercise resulting in more positive filling pressures, teleosts may further shift to a vis-a-tergo cardiac filling mode for increased activity (85). However, the relative proportion of early to late phase filling (passive vs active atrial filling) suggests that the late phase is
greater than the early phase in biphasic (65) and the third phase in triphasic does not change this pattern. With late phase (active atrial contraction) playing such an important role regardless of the filling mode, the trend still holds that fish hearts are still more dependent on atrial contraction for ventricular filling than mammals. This results in species such as zebrafish having relatively low systolic pressure generation (~2.5 mmHg) in the ventricle (51).

3.3. Force generation

Cardiac performance may be modulated through variations in stroke volume and heart rate, but the relative contributions vary substantially among fish species. In general fish respond to greater hemodynamic loads by increasing cardiac output mainly through increased stroke volume rather than heart rate. Stroke volume can vary up to 2-3 fold with exercise but only in specific species, such as rainbow trout and kingfish (17, 18). End systolic volume (ESV) of the fish heart is very small (20) unlike that of the mammalian heart. This results in lower pressure in the teleost ventricle as a whole (51). The end systolic stroke volume in the fish ventricle is greatly determined by atrial contraction rather than central venous pressure seen in mammals (19). Increased importance of late filling created by atrial contraction rather than the passive early filling of the ventricle is reflected in the early to late filling ratio (E/A ratio) of teleosts. Fish typically show an E/A ratio around 0.2, much lower than the typical healthy human (41). The modulation of atrial contractility and Frank-Starling mechanisms are thus particularly important in determining ventricle filling.

The whole heart displays a greater sensitivity to filling pressure than in mammals, in part attributed to greater myocardial extensibility coupled with an increase in myofilament Ca$^{2+}$ sensitivity over a large range of sarcomere lengths (101). The Frank-Starling principle dictates that the strength of contraction is, in part, a function of the length of the muscle fiber (and hence sarcomere length) prior to contraction (i.e. preload). Stroke volume increases with end diastolic volume since muscle fibers and sarcomeres within the fibers are stretched towards their optimum. Between species of fish there is
considerable variation in length-dependent sensitivity. For example, flounder heart is more sensitive to filling pressure reflecting the high distensibility of its chambers (71). This length-dependent contraction strength is also directly affected by temperature in certain teleost species (e.g. trout- (101)). With fish cardiac myocytes showing enhanced length-dependent activation, and Ca$^{2+}$ sensitivity increasing with sarcomere length (80), Ca$^{2+}$ handling is an important consideration in the basis of the length dependence of the strength of cardiac contraction. The reduced relative duration of diastole seen with active atrial contraction in fish may have evolved in response to the relatively slow rate of the excitation-contraction coupling processes.

The importance of strong atrial contraction is even seen with atrial muscle having the ability to generate greater maximal force ($F_{\text{max}}$) than ventricular muscle in a variety of teleost species (e.g. rainbow trout- (2); yellowfin tuna - (96)). Although systolic blood pressure measurements show the atrium to be a weaker pump, the higher force suggests the relative weakness of the atrium may be reflective of the anatomical difference in wall thickness between the chambers. Greater numbers of cardiomyocytes in thicker ventricle walls generate greater overall force, while greater $F_{\text{max}}$ of individual atrial cardiomyocytes can be generated in excised strips (94).

Cardiac performance is directly related to the contractile properties of cardiomyocytes. Ventricular and atrial muscles express different paralogs (or isoforms) of myosin heavy chain (MHC) (59) and this predominantly determines the maximum force produced by muscle fibers. Alterations in paralog composition therefore are associated with changes in myofibrillar function as well as myosin ATPase activity (9). Because of this, the myosin ATPase reaction is higher in atrial than in ventricular cardiomyocytes (60). Across vertebrates, from the primitive dogfish up to mammals, the atrium has a higher total myosin heavy chain (MHC) protein expression than the ventricle (28). The increased levels of MHC may be contributing to the increased $F_{\text{max}}$ of individual atrial cardiomyocytes relative to ventricular cardiomyocytes. However, MHC expression pattern differences between chambers may also be species-specific. In goldfish, both chambers have electrophoretically-similar MHC while in other fish species such as pike, ventricular myocardium exhibited electrophoretically faster MHC than seen in the atrium (60). Linking chamber-specific isoform usage with functional differences across a broader phylogenetic
range as well as temperature variations might help to clarify the importance of this heterogeneity.

### 3.4. Action potential

Cardiac performance is directly related to the electrophysiological properties of the heart. The morphology of the heart sets up an alternating arrangement of slow conducting segments (e.g. inflow tract, atrioventricular region and outflow tract) with fast conducting contractile segments (atrium and ventricle) (75). Even with all the morphological differences between the two-chamber teleost heart and the mammalian design, the electrophysiological properties are similar (11, 67). Whole heart measurement in zebrafish demonstrates intrinsic heart rate (HR) variability with the excitation originating near the atrial border, opposite the AV junction (67). This region corresponds to a ring shaped grouping of cells expressing *Isl1* and *Tbx2b* at the junction of the sinus venosus and the atrium and has been both molecularly and structurally linked to pacemaker activity (109).

The ventricular AP is initiated by a voltage-gated increase in Na⁺ channel conductance resulting in a greater sodium current (INa⁺). Repolarization is generated by an increase in K⁺ conductance that is slower to develop but longer lasting (IK⁺). The plateau of partial repolarization is due to the persistence of IK1 and activation of L-type Ca²⁺ channels creating an inward Ca²⁺ current that is relatively slow to inactivate. Sarcolemmal (SL) Na⁺/K⁺ pumps restore intracellular Na⁺ and K⁺ concentrations while the sarcolemmal Na⁺/Ca²⁺ exchanger (NCX) and SR/SL Ca²⁺ ATPases restore the cytosolic Ca²⁺ concentration to diastolic levels (116).

The excitability of myocytes and conduction velocity of cardiac AP are primarily determined by INa⁺. The resting membrane potential of atrial cardiomyocytes is lower than that of ventricular cardiomyocytes (approximately -40 mV in atrial vs -70 mV in ventricular enzymatically isolated cells) (117). IK1 is the predominant current in ventricular cells while IK is predominant in atrial cells (117). These variations in K⁺ current densities appear to cause the atrial AP duration to be much shorter than ventricular AP duration (e.g. carp - (118); trout - (117); zebrafish - (77)). This chamber-specific variation in K⁺ currents, along
with the more negative voltage threshold of $I_{Na}$ and higher input resistance in atrial cardiomyocytes lead to higher excitability in the atria (trout - (40)). This allows pacemaker cells to elicit atrial excitation while being influenced themselves by atrial APs, while ventricular cardiomyocytes can have higher thresholds for excitation to avoid accidental firing. Regardless of $I_{Ca2+}$, the variations in AP shape and duration brought on by K+ currents may be significant in affecting SL Ca2+ influx, since the shorter AP plateau seen in the atrium provides less time for extracellular Ca2+ influx to occur (77).

3.5. Excitation-contraction coupling

Excitation-contraction coupling involves converting the electrical signal of an action potential to the mechanical response of contraction. The depolarization of an action potential causes the opening of L-type Ca2+ channels (CaV1.2) which are responsible, at least in part, for a cytosolic Ca2+ transient that then activates contractile proteins. Hence, Ca2+ handling is an important consideration as the primary basis for regulation of cardiac contraction in all species, including fish. The rate of diffusion of extracellular Ca2+ to and from cardiac troponin C (cTnC), the thin filament protein which senses changes in cytosolic Ca2+ concentrations and initiates the contractile process, sets the limit for contraction frequency. The basic features of Ca2+ release through the ryanodine receptor are conserved from teleosts to mammals (10, 70). However, variation is seen in Ca2+ sources, myofilament Ca2+ sensitivities and/or spatial arrangement of the contractile element to function in the environmental conditions teleosts exist in. Mammalian cardiac muscle has intracellular Ca2+ stores in the SR in close proximity to t-tubules that provide the majority of the Ca2+ that binds to cTnC. The contribution of SR Ca2+ stores to contractile activation has traditionally been seen as minimal in endothermic species (117) with the majority of Ca2+ derived from extracellular space (24, 101, 110, 117) despite large SR Ca2+ loads (50, 70, 120). While the degree of Ca2+ derived from extracellular space is quite variable between teleost species, this pattern reflects the lack of t-tubules seen in ectothermic myocytes. The measurement of Ca2+ cycling has been important in comparisons of contractile properties of teleost cardiomyocytes relative to both mammals and between teleost species.
3.5.1. Activation of Contraction

The comparison of teleost with mammalian cardiomyocytes in terms of the key contributors to the activation of contraction is difficult due to variability seen across teleosts. Key contributors of Ca\textsuperscript{2+} in mammals are the L-type Ca\textsuperscript{2+} channels that mediate Ca\textsuperscript{2+} entry across the sarcolemma and SR Ca\textsuperscript{2+} release channels (RyR2), which release the majority of the Ca\textsuperscript{2+} from the SR. The action potential first initiates Ca\textsuperscript{2+} release via the L-type Ca channel (I\textsubscript{CaL}). This further initiates Ca\textsuperscript{2+}-induced Ca\textsuperscript{2+} release (CICR) via the SR. This release is measured as Ca\textsuperscript{2+} sparks initiated with release of Ca\textsuperscript{2+} from the SR, which leads to synchronized Ca\textsuperscript{2+} release from the SR via a large number of ryanodine receptor clusters. Hence in mammals, CICR plays a significant role in activating contraction. During the upstroke of the action potential, reverse mode NCX may also contribute (7).

Since the initial reports on I\textsubscript{CaL} measurements in fish cardiomyocytes (47, 110, 113) there is now substantial information on I\textsubscript{CaL} density in atrial and ventricular myocytes from a number of fish species under different experimental conditions (98, 99). In teleosts, differences in both experimental conditions and species are abundant and there are substantial differences in the measurements of Ca\textsuperscript{2+} (47, 111). Estimations of the contribution of the Ca\textsuperscript{2+} carried by I\textsubscript{Ca} to the activation of contraction also varies from as little as 25-30% in trout atrial myocytes using a perforated patch technique (46, 47), to more than 50% in zebrafish ventricular myocytes subjected to perforated patch-clamp technique (120) and close to 100% using ruptured patch-clamp technique carp ventricular myocytes (113) and in trout atrial myocytes (44). This does not necessarily imply that I\textsubscript{Ca} is the sole source of Ca\textsuperscript{2+} delivery in these preparations but merely indicates that it is sufficiently large to activate contraction if no other Ca\textsuperscript{2+} sources are available. Chamber-specific differences in the I\textsubscript{Ca} density have not been systematically investigated in the teleost heart.

Reverse mode NCX can also contribute to the activation of contraction at depolarized membrane potentials and/or when sodium channels produce a large local increase in the subsarcolemmal Na\textsuperscript{+} concentration. Several studies have shown that fish cardiomyocytes have a high NCX activity relative to mammalian cardiomyocytes, and the
direct contribution of the NCX to the activation of contraction has been estimated to range from 25-45% under physiological conditions (44, 46, 120). This NCX contribution could even be equal to that of the L-type Ca$^{2+}$ current (44, 45). However, the relative contribution of $I_{\text{NCX}}$ to total SL Ca$^{2+}$ entry depends critically on the membrane potential and the intracellular Na$^+$ concentration (40, 46, 119). Moreover, there is evidence that reverse mode NCX can trigger Ca$^{2+}$ release from the SR in trout atrial myocytes (44).

Ca$^{2+}$ release from the SR has also been reported to contribute to the activation of contraction in fish cardiac myocytes. SR contributions vary between zero to more than 50% depending on species, tissue and experimental temperature. Ryanodine, a compound that modulates the opening of the SR Ca$^{2+}$ release channel, was first reported to have marginal effects on contraction in multicellular tissue preparations at physiological temperatures and stimulation frequencies (23, 33). Ryanodine was however reported to inhibit post rest potentiation, a phenomenon caused by Ca$^{2+}$ release from the SR (43). Ryanodine has since been reported to have a strong effect on contraction in trout (42) and skipjack tuna (61) at 25ºC, suggesting a temperature-, dose and/or species-dependent component of the effect of ryanodine.

In fish, both $I_{\text{Ca}}$ and reverse mode NCX are capable of triggering Ca$^{2+}$ release from the SR (44). The relative contribution of the two mechanisms depended strongly on the membrane potential and the intracellular Na$^+$ concentration (8). Steady-state and maximal SR Ca$^{2+}$ content in trout ventricular myocytes were measured at about 10 times higher than values reported in human cardiomyocytes. The SR uptake rates were sufficiently fast for the SR to participate in the regulation of Ca$^{2+}$ handling on a beat-to-beat basis (50). Teleost cardiomyocytes have high SR activity even though sufficient Ca$^{2+}$ is provided through L-type current channels required for activation of contraction (70, 120). SR Ca$^{2+}$ is more likely to be released in trout atrial than ventricular myocytes, based on the frequency of spontaneous membrane depolarizations induced by SR Ca$^{2+}$ release (14). The loss-of-function NCX mutant zebrafish (tremblor or tre mutant) demonstrates more cardiac fibrillation in the atrium relative to the ventricle (66), further evidence of the atrium being more susceptible to spontaneous Ca$^{2+}$ release from the SR as a result of disruptions of normal Ca$^{2+}$ transients.
3.5.2. Relaxation of Contraction

The Ca\(^{2+}\) delivered to the myofilaments during the activation of contraction has to be removed from the cytosol during relaxation. Moreover, if equilibrium is to be maintained, the Ca\(^{2+}\) entering across the SL must be extruded from the cell and Ca\(^{2+}\) released from intracellular organelles must be re-sequestered by the same organelle. In both mammalian and trout hearts, forward mode NCX generally accounts for the vast majority of the SL Ca\(^{2+}\) extrusion (5)(47)(50). Ca\(^{2+}\) extrusion by NCX in trout ventricular myocytes was estimated to account for about 55% of the total Ca\(^{2+}\) removed during relaxation (46).

SR Ca\(^{2+}\) content in trout ventricular and atrial myocytes can be up to 10 times higher than in mammalian myocytes (47, 96, 98, 101). Sarcoendoplasmic reticulum Ca\(^{2+}\) ATPase (SERCA) plays a critical role in the recycling of cytosolic Ca\(^{2+}\) back into the SR that contributes to the relaxation of cardiac muscle. Variation in rates of Ca\(^{2+}\) recycling cannot be only be attributed to the differences in SERCA, since SERCA’s activity is tightly regulated by phospholamban (PLN). In its dephosphorylated state, PLN monomers bind to SERCA and inhibit SR Ca\(^{2+}\) pump activity. Inhibition is relieved under beta adrenergic activation, with PKA and/or Ca\(^{2+}\)/calmodulin-dependent protein kinases phosphorylation of PLN result in increased Ca\(^{2+}\) affinity of the SERCA and faster SR Ca\(^{2+}\)-uptake (69). Differences also appear between the chambers in the degree of contribution from SR Ca\(^{2+}\) stores. It is interesting to note that the higher expression of the SERCA pump is thought to provide a faster rate of Ca\(^{2+}\) sequestration in atrial vs. ventricular muscle of trout (65). SR Ca\(^{2+}\) load in fish atrial and ventricular cardiomyocytes has been measured as high enough that only 10% of the total SR Ca\(^{2+}\) load would be released on each beat (45, 47, 100). The rate of SR Ca\(^{2+}\) uptake under physiological conditions was also sufficient to remove the total Ca\(^{2+}\) required for the activation of contraction between beats at physiological beating rates and temperatures in trout atrial and ventricular myocytes (49, 50). The reason for such an excessive SR Ca\(^{2+}\) load compared to the Ca\(^{2+}\) requirements on a beat-to-beat basis remains elusive. Nonetheless, it definitely provides fish species with an intracellular Ca\(^{2+}\) reservoir that allows the heart to function under adverse conditions where SR Ca\(^{2+}\) refilling is impaired for extended periods of time.
Differences in chamber-specific expression of the contractile element proteins responsible for Ca\(^{2+}\) sensitivity may also guide differences between contractility of the atrium and ventricle. These changes in the regulation of cardiac contractility occur through alteration of cardiac-specific isoforms of sarcomeric proteins like troponin C (78, 104), troponin I (3), tropomyosin (121), troponin T (93), myosin light chain (83), myosin heavy chain (60) and titin (92). With lower L-type Ca\(^{2+}\) current densities, higher Ca\(^{2+}\) sensitivity of the myofilament (34, 36) may be necessary for contraction in teleost hearts as well as additional reliance on alternative Ca\(^{2+}\) sources. Myofilament Ca\(^{2+}\) sensitivity is a critical factor in regulating cardiac contractility. Zebrafish cardiac TnT has been shown to be required for the contraction of both the atrium and the ventricle (93) similar to mammals. However, zebrafish possess two cTnC gene copies, with one predominantly expressed in the embryonic zebrafish heart. Loss of function of this gene results in abnormal atrial and ventricular chamber morphology, but the second cTnC gene enables function to be regained in the atrium (105). This suggests a possible difference in expression of cTnC between chambers that could potentially lead to variation in Ca\(^{2+}\) sensitivity. Gene duplication has led to multiple copies of sarcomeric proteins, and variation between these fish-specific paralogs may be a method of coping with varying environmental conditions.

3.6. Variation in cardiac contraction with environmental perturbation

The interspecific variation of fish cardiac anatomy and physiology is guided by evolutionary adaptation to different habitats, modes of life and activity levels. Because ectothermic fish are exposed to different temperatures either seasonally or via the vertical thermocline, temperature-dependent regulation of cardiac contractility is crucial. Many fish face unpredictable temperature changes, meaning that fish hearts must have intrinsic compensatory mechanisms or rely on post-translational modifications to protect against acute temperature changes when there is not enough time for transcriptional regulation to take effect (111, 117). Most changes that occur with acute temperature fluctuation are readily reversible. Only certain species of fish have the ability to tolerate significant acute temperature changes (106), typically eurythermal species with an evolutionary history of
exposure. Acclimation, or remodeling to cope with longer-term changes in temperature, typically involves transcriptional regulation, which results in morphological or biochemical changes in order to cope with the new environment. Temperature adaptation is one example of an environmental stressor that can modulate contractility and is a crucial component of this thesis.

Cardiac performance may change through independent changes in stroke volume and heart rate, with relative contributions of each varying substantially among fish. During exercise, more athletic fish modulate cardiac output preferentially by increasing SV, while different species preferentially modulate HR in different species (17, 18). With environmental hypoxia, reflex bradycardia occurs and increased stroke volume is necessary to maintain cardiac output (30, 81). Fish typically exhibit changes in HR with acute temperature fluctuation (16, 37) but the extent to which this varies cardiac output depends on the species. Higher temperatures have been shown to increase heart rate while colder temperatures decrease stroke volume in crucian carp (115). Longer term decreased acclimation temperature in trout leads to increases in overall size of the ventricle as well as connective tissue content resulting in increased stroke volume (62). While acute temperature change modulates HR in zebrafish (67), cold acclimation can increase maximal metabolic rate but does not affect resting HR (68), or heart volume (58). Zebrafish hearts may maintain cardiac output with decreased temperature by reducing the thickness of the compact myocardium and modulating collagen content to maintain compliance of the myocardium (58). Variation in modulation of cardiac performance to short-term and long-term temperature change does appear to be species-specific and must be quantified accurately.

A limiting factor in the ability to cope with acute changes in temperature is adequate cardiac contractility, so species must possess sufficient myocardial plasticity (79). In teleosts, the need for higher relative Ca$^{2+}$ sensitivity for cold temperatures must be balanced with the need for greater variability due to changing environmental conditions. The large SR Ca$^{2+}$ stores and low Ca$^{2+}$ sensitivity of ryanodine receptors have been proposed to play a critical role in maintaining contraction at lower temperatures than mammals (95, 97). Higher SR Ca$^{2+}$ content may be necessary to initiate and maintain SR Ca$^{2+}$ release events at lower temperatures (38, 48). Interspecies variation in excitation-
contraction coupling seen with temperature acclimation may be indicative of the capacity for thermal niche expansion. This substantial variation in SR Ca\textsuperscript{2+} release relative to SL Ca\textsuperscript{2+} influx ratio exists between species as well as being dependent on acute and chronic temperature changes. Cold-tolerant active fish have a larger capacity to store Ca\textsuperscript{2+} in the SR than mammals (e.g. active teleosts - (38); scombroids - (29)). The atrioventricular differences in Ca\textsuperscript{2+} loading are also more pronounced in cold-acclimated than warm acclimated fish (39). The degree of modification of I\textsubscript{Ca} to offset the negative effects of colder temperatures appears to be species specific, with many teleosts such as trout demonstrating relatively temperature-insensitive Ca\textsuperscript{2+} flux (100) whereas certain scombroids show significant reductions in I\textsubscript{Ca} with decreasing temperatures (39). The expression of the SERCA is higher in atrium than ventricle of the trout heart and a lower PLN/SERCA ratio is also seen in cold-acclimated trout hearts in comparison to warm-acclimated trout hearts (63). This is thought to contribute to chamber- and temperature-related differences in SR Ca\textsuperscript{2+} cycling. Due to the whole genome duplication in the teleost lineage, addition paralogs of SERCA genes and their protein products exist in fish relative to tetrapod vertebrates. The expression of SERCA transcripts does not appear to follow the expression pattern of SERCA protein, suggesting that SERCA expression has a strong post-transcriptionally regulation component in trout (64). However, this does provide an example of the impact of whole genome duplication on contractile properties of the fish heart.

The presence of multiple fish-specific paralogs of important components of the contractile element may be indicative of how fish have evolved mechanisms to thrive in variable environments. Incorporating the chamber specific variation with temperature response reveals the complexity of these mechanisms in achieving survival. Differing paralog usage between the chambers may be linked to temperature-induced Ca\textsuperscript{2+} sensitivity differences. Multiple paralogs of TnI are expressed in cardiac tissue. The relative expression of these paralogs in the ventricle shifts with temperature, possibly leading to changes in the Ca\textsuperscript{2+} affinity of the entire troponin complex (3). In fish, multiple TnC paralogs are present in cardiac tissue (105). Variations between chambers of Tn subunits may also lend insight into Ca\textsuperscript{2+} sensitivity changes with temperature, as the affinity of cTnC in trout is temperature dependent (35). Modulation of MHC isoform expression patterns with temperature also leads to changes in myofibrillar ATPase activity.
Ventricular and atrial muscles express different paralogs (or isoforms) of MHC (60). These isoforms are not only divided spatially but also by differences in timing of expression. Atrial-specific myosin gene expression is induced after ventricle-specific myosin gene expression in zebrafish (6). Modulation of contractile properties via longer-term transcriptional changes is not always possible, therefore plasticity due to mixed expression of paralogs for critical components of contractile unit may be an important facet of the ability of fish to cope with environmental change. Fish-specific genome duplications allows for multiple chamber-specific genes with variable functions that allow for flexibility in function. Further work is required to determine how genome duplication has shaped the structure and function of the fish heart and to clarify how this wide-range of species-specific phenotypes has been maintained.

3.7. Summary

The unique structure of the fish heart accommodates the variability in function allowing species to exploit many different environments. The increased importance of the atrial contraction in guiding changes in cardiac output indicates that some of this physiological versatility may be due to atrio-ventricular differences. These two main chambers of the fish heart differ in terms of contractile properties from the molecular level of E-C coupling and AP morphology, and the whole heart morphology and function reflect these differences. The fish heart exhibits, in some ways, similarities to the early embryological stages of higher vertebrates, hence many studies on both the development and functioning of chamber-specific properties lends insight into proper functioning of the heart not only in fish. However, the degree of atrio-ventricular differences vary between species making it very difficult to generalize species-specific results across phylogeny. The species-specific differences in these two chambers lend insight into how evolutionary history guides responses to environmental factors such as ambient temperature. The characterization of functional properties of the heart are critical for the exploitation of the adult zebrafish as a model for cardiac function, as well as for understanding of the responses of teleost cardiac function to environmental perturbations.
3.8. References


24. Fabiato A. Calcium-induced release of Ca2+ from the cardiac sarcoplasmic reticulum. Am J Physiol 245: C1-14, 1983.


Chapter 4.
Functional assessment of cardiac responses of adult zebrafish (*Danio rerio*) to acute and chronic temperature change using high-resolution echocardiography

This chapter is largely based on the following original research article:


‡ These authors contributed equally to this work.

My contribution to this study included experimental design, acclimation of fish and analysis of the data set. I was responsible for writing the manuscript.
4.1. Abstract

The zebrafish (*Danio rerio*) is an important organism as a model for understanding vertebrate cardiovascular development. However, little is known about adult ZF cardiac function and how contractile function changes to cope with fluctuations in ambient temperature. The goals of this study were to: 1) determine if high resolution echocardiography (HRE) in the presence of reduced cardiodepressant anesthetics could be used to accurately investigate the structural and functional properties of the ZF heart and 2) if the effect of ambient temperature changes both acutely and chronically could be determined non-invasively using HRE *in vivo*. Heart rate (HR) appears to be the critical factor in modifying cardiac output (CO) with ambient temperature fluctuation as it increases from 78 ± 5.9 bpm at 18°C to 162 ± 9.7 bpm at 28°C regardless of acclimation state (cold acclimated CA – 18°C; warm acclimated WA – 28°C). Stroke volume (SV) is highest when the ambient temperature matches the acclimation temperature, though this difference did not constitute a significant effect (CA 1.17 ± 0.15 µL at 18°C vs 1.06 ± 0.14 µL at 28°C; WA 1.10 ± 0.13 µL at 18°C vs 1.12 ± 0.12 µL at 28°C). The isovolumetric contraction time (IVCT) was significantly shorter in CA fish at 18°C. The CA group showed improved systolic function at 18°C in comparison to the WA group with significant increases in both ejection fraction and fractional shortening and decreases in IVCT. The decreased early peak (E) velocity and early peak velocity / atrial peak velocity (E/A) ratio in the CA group are likely associated with increased reliance on atrial contraction for ventricular filling. Remodeling of atrial usage with cold acclimation in zebrafish is an important response in varying cardiac contractility with environmental challenge.
4.2. Introduction

Ectothermic fish must cope with environmental temperature changes on a chronic (seasonal) and on an acute basis. Changes in water temperature can have drastic effects on both resting and maximum cardiac performance (19, 43). Only certain species of fish have the ability to tolerate significant acute temperature changes (42), typically eurythermal species with an evolutionary history of exposure (either on a daily basis and or via water column movements). A limiting factor in the ability to cope with these acute changes is adequate cardiac contractility, so species must possess sufficient myocardial plasticity (38). Fish typically exhibit changes in heart rate with acute temperature fluctuation (6, 19) but the extent to which this varies cardiac output depends on the species. Most changes that occur with acute temperature fluctuation are readily reversible. Acclimation, or remodeling to cope with longer term changes in temperature, typically involves transcriptional regulation which results in morphological or biochemical changes in order to cope with the new environment. Since these changes take longer to accomplish, there is often a cost associated with them, such as the energetic cost of producing proteins or an expression profile that is no longer optimal for the current environment.

While temperature effects on metabolic processes have been well documented, the overall morphological and functional changes to the adult teleost heart have not been as comprehensively studied. The teleost heart is composed of a series of four chambers: venous sinus; atrium; ventricle and bulbus arteriosus. These chambers develop from a simple linear tube (35) and differ not only morphologically but also physiologically with different contractility characteristics. Compared to mammals, the relative volume of the fish atrium is closer to that of the ventricle and despite being less developed muscually than the ventricle, the teleost atrium plays a much more important role in the active filling of the ventricle. In both mammals and fish a key regulator of cardiac performance is the end diastolic volume (EDV). EDV in fish is greatly determined by active atrial contraction rather than by central venous pressure as seen in mammals (7). Thus the modulation of atrial contractility as well as Frank-Starling mechanisms involving preload in the ventricle are particularly important for determining ventricle filling in teleosts. The reported degree
to which the active role of the atrium has modulated cardiac output in the literature has varied depending on both the experimental set-up and the species (3, 4, 16, 24). Many of the morphological differences between chambers may have evolved in order to meet the high physiological demands of variation in metabolic rate and increased tolerance to a wide range of environmental conditions. Chamber-specific properties may lend themselves to teleost-specific functional responses to ambient temperature. Subjecting the heart to acute temperature stress allows one to look at the intrinsic ability of the fish heart to cope with rapid changes in temperature that do not allow for transcriptional level remodeling. Conversely, longer-term acclimation to temperature stress allows us to examine more extensive remodeling of the heart guided by changes in gene expression profiles. Both short-term plasticity and longer-term remodeling are important components constituting the complex response of teleosts to environmental challenges.

Many previous studies have used a reductionist approach to determine the potential limitations of heart performance with respect to temperature; however, the variability in results due to experimental set-up makes it difficult to determine the \textit{in vivo} effects on cardiac contraction. Conventionally, cardiac hemodynamics have been measured by invasive catheterization techniques in fish (14, 16, 20) but these involve manipulation of the \textit{in situ} environment to mimic \textit{in vivo} conditions leading to inconsistencies in the background. Stroke volume (SV) may be derived from cardiac output, leading to underestimations of true SV (16). Doppler technology has been used to study intracardiac flows non-invasively, but its use in fish has been largely limited to surgically implanted probes in some cases requiring potential error-introducing calibration after removal from the animal (44), as well as being limited to fish larger than zebrafish.

The use of echocardiography with a high frequency (50 - 70 MHz) probe can allow for high resolution, real-time, noninvasive imaging to examine many parameters of cardiac structure and function. The zebrafish has been proposed by some to be a robust model of human cardiac electrophysiology due to the fact that its heart rate and action potential morphology are similar although some differences are apparent (36, 49). Zebrafish have a wide seasonal temperature range from 6°C to 38°C and its high temperature tolerance make this real-time technique ideal for demonstrating how the cardiac function of eurythermal species cope with temperature fluctuation (41). The use of conventional
ultrasound equipment (7 and 8.5 MHz) in previous studies on zebrafish (20) had insufficient resolution leading to problems with visualization of specific chambers and specific valves, or conventional use of relatively high doses (100-200 ppm) of the fish anesthetic MS-222 (46) confounds measurements of HR and CO through its well-documented cardio-depressive effects (20, 46). In this study we used high-resolution echocardiography using a more appropriate anesthetic to look at the structural and functional responses of the zebrafish heart to ambient temperature changes as a result of acute change and longer-term acclimation. We clearly demonstrate the ability of high-resolution echocardiography to accurately and non-invasively determine multiple parameters of cardiac structure and function in vivo in the adult zebrafish and firmly establish the utility of this technique in characterizing responses to environmental perturbations in similar-sized teleosts.

4.3. Methods

4.3.1. Animals and cold acclimation

All animal experiments were approved by the University of British Columbia Animal Care Committee (protocol number A12-0198) and adhered to the Canadian Council on Animal Care (CCAC) guidelines. Warm-acclimated (WA) adult zebrafish purchased from a local supplier were maintained at a 12-h:12-h light:dark photoperiod at 28°C in dechlorinated water in a flow-through tank and fed ad libitum (Nutrafin Max, Hagen Inc., Baie d’Urfé, QC, Canada). Based on size (48), all fish were estimated to be at least 20 weeks of age at the start of the experiment. The cold-acclimated (CA) group of adult zebrafish was held at 28°C for two weeks in flow-through tanks, followed by a temperature decrease of 2.5°C/week until the end-point temperature of 18°C was reached. Fish were then held at this temperature for at least 4 weeks.
4.3.2. Zebrafish holding conditions

For echocardiography, WA fish were held for 7 days in an aquarium in which the temperature was maintained at 28 °C by means of an Eheim Jäger aquarium thermostatic heater (Deizisau, Germany). CA fish were also held for 7 days in a chilled aquarium with the temperature maintained at 18ºC. The aquarium water was double distilled and a GH Mineralized mineral supplement was added (Aquavitro, Madison GA) in a ratio of 1 ml / 8 L to maintain minimum calcium and magnesium concentrations of 13.5% and 1.2%, respectively. The pH was adjusted to 7.2 with NaOH.

4.3.3. Anaesthetic preparation

MS-222 (Tricaine methane sulfonate) is used extensively as an anesthetic in fish studies, however at the concentrations typically used (100 – 200 ppm or mg/L), it acts as a profound cardiodepressant (22) and can seriously confound the results. In this study a combination of 45 ppm of pH-adjusted MS-222 (Sigma-Aldrich, Oakville, ON) and 45 ppm isoflurane (Sigma-Aldrich, Oakville, ON) was used for zebrafish to allow anaesthesia time to be extended with less consequence to cardiac function in comparison to MS-222 alone (23). To create stock solutions, MS-222 was dissolved in distilled water to a final concentration of 2,000 ppm and buffered to pH 7.4 with NaOH and isoflurane was dissolved in 100% ethanol in a ratio of 1:9 to a final concentration of 100,000 ppm. Zebrafish were placed in a 45 ml custom-made water-jacketed glass chamber containing 30 ml of aquarium water and an initial concentration of 30 ppm of each drug. After 15 minutes, a second dose of MS-222 and isoflurane anesthetic mix (15 ppm) was added to yield final concentrations of 45 ppm of each drug. It was determined that the appropriate anesthetised condition for echocardiography was reached when the zebrafish were unresponsive to a slight tail pinch with forceps. Breathing was monitored by visual tracking of movement of the opercula to ensure fish health throughout the protocol.
4.3.4. Echocardiography

Anesthetized adult zebrafish (n=10 for each treatment group) were held by a small piece of modelling clay (Play-Doh™ Hasbro, Longueuil, QC) on the tail and placed ventral side up in the glass chamber. The water jacket was connected to a water bath to ensure precise temperature control. Aquarium water temperature was monitored by a temperature probe submerged in the bath chamber. Echocardiographs of the zebrafish were acquired by using a Vevo 2100 ultrasound system (VisualSonics®, Toronto, ON, Canada), with a 70 MHz ultrasound transducer fixed above the ventral side of the zebrafish and parallel to the longitudinal axis plane (Figure 4-1). The frame rate ranged from 20 to 120 (depending on image size) frames per second. The specifications of the transducer (Model MS700) used in this study included having a central frequency of 50 MHz, bandwidth from 30-70 MHz, a focal length of 5.0 mm and a maximum frame acquisition rate of 476 fps (single zone, 4.73 mm width, B-mode). The maximum field of view of 2D imaging was 9.7 x 12.0 mm with a spatial resolution of 75 µm (lateral) by 30 µm (axial). All calculations were done using the cardiac package of the VisualSonics software.

Figure 4-1  Longitudinal position of zebrafish. The ultrasound transducer beam (70 MHz) was positioned at about 85 degrees to ventral surface of the anesthetized zebrafish.
Cardiac functional parameters such as SV, ejection fraction (EF), fractional shorting (FS) and cardiac output (CO) were obtained from a clear two chamber B-mode image (Figure 4-2). Ventricular volume was determined from the differences between systolic and diastolic ventricular areas. This was measured by the rotational volume of the ventricular trace around the long axis line of the spline, referring to the long axis line length that extends from the base to the farthest extent of the spline through systole (Ls) and diastole (Ld) (Figure 4-2). Fractional shortening (FS) was calculated by (Ld - Ls) / Ld multiplied by 100. Stroke volume (SV) was measured by the difference between diastolic volume and systolic volume, while ejection fraction (EF) was measured by SV divided by diastolic volume multiplied by 100 (Fig 4-2). The color Doppler and pulsed wave Doppler images were also captured from the same longitudinal plane. Heart rate, ventricular inflow and ventricular outflow peak velocity were calculated from the pulse wave Doppler images (Figure 4-3). These measurements were acquired at the point between the atrium and the ventricle for inflow velocity, and between the ventricle and the bulbous arteriosus for outflow velocity. The direction of blood flow was determined using color Doppler, meaning the angle used for quantification of peak velocity can be measured precisely in each fish. The Doppler velocity of the ventricular inflow includes two components: early (E) filling peak velocity that occurs during ventricular relaxation and the atrial (A) filling peak velocity that is the consequence of atrial contraction. Isovolumetric relaxation time (IVRT), isovolumetric contraction time (IVCT) and ejection time (ET) (measured in ms) were also collected from pulsed wave Doppler images of ventricular inflow (Figure 3). IVRT was taken pre-early (E) filling peak and IVCT was taken post-atrial (A) filling peak. MPI was calculated from (IVCT+IVRT)/ET.

Every parameter was determined as the average measurement of at least five cardiac cycles. Each acclimated fish was first measured at their respective acclimated temperature (i.e. 18°C for CA and 28°C for WA). After the first image was acquired, the temperature of the chamber was gradually changed over the course of twelve minutes from 28 to 18°C for WA, and then back to 28°C over the same time span. The same procedure was followed in reverse for CA. The ultrasonic images were taken in a stepwise manner at temperature points of 18 and 28°C. The temperature was maintained for 2 minutes before echocardiographic images were acquired to ensure adequate cooling or warming of the zebrafish body. For three of the 20 zebrafish used in this study, an
additional 7.5 ppm of both MS-222 and isoflurane were added in order to maintain the anesthetic condition while the temperature was raised to 28°C without any apparent cardiodepressant effects.

**Figure 4-2**  B-mode image of a long-axis view of a WA zebrafish heart at 28°C.

The trace indicates the perimeter of the ventricle represented in A) red for Diastole and B) cyan for Systole. VA, ventricular area; A, area; V, volume; Ld, greatest length of the spline during diastole; Ls, greatest length of the spline during systole; SV, stroke volume determined by diastolic volume-systolic volume; EF, ejection fraction determined by SV/diastolic volume; FS, fractional shortening determined by (Ld - Ls) / Ld * 100.
4.3.5. Statistics

Statistical significance in acute temperature responses for echo parameters were determined using a one-factor two way ANOVA with a randomized block design. The acclimation state was treated as a blocking variable allowing the potential interaction of...
this effect with temperature to be explored. If there was no interaction, the statistical analysis was redone without the interaction term. If the interaction was significant, acclimation state has a confounding effect which cannot be dissociated from the effect temperature change on the response variable. Values outside of two standard deviations from the mean were discarded from the analysis. A $p$-value of less than 0.05 was accepted as statistically significant. The JMP 11 software package was used for all statistical analysis.

### 4.4. Results

#### 4.4.1. Functional Analyses

Heart rate was used as the principle parameter to assess whether the anesthetic protocol caused significant cardiac depression that would confound the accurate determination of cardiac function. The combination of MS-222 and isoflurane kept the heart rate above 160 bpm in fish at 28°C. An acute temperature decrease of 10°C significantly decreased heart rate by approximately half ($p<0.0001$, Figure 4-4a). Cold acclimation did not significantly modify the acute temperature heart rate response, as there was no significant difference in heart rates taken at both 28°C (162-169 bpm) and 18°C (78-79 bpm) for both CA and WA fish. SV did not fluctuate significantly with acute temperature perturbation and cold acclimation had no statistically significant effect on this response (1.06 - 1.17 μl). Combining these two factors, CO was significantly increased by about two-fold for both groups in increasing the acute temperature from 18 to 28°C ($p<0.0001$) and acclimation state had no significant effect on the magnitude of CO at either temperature (Figure 4-4b).

Ejection fraction, the volumetric fraction of blood pumped out of the ventricle with each cardiac cycle relative to the maximum volume in the heart after diastole (EDV), can be used as an index of contractility. Acute temperature perturbation did not modify EF significantly in both groups, but acclimation state had an interacting effect on the temperature response (cross-effect $p = 0.018$). With CA, EF was higher at both
temperatures (18 and 28ºC) and increased when temperature was decreased to 18ºC (Figure 4-4c).

Figure 4-4  Ventricular functional parameters were calculated from B-mode long-axis plane (n = 10).
A) Heart rate, B) normalized cardiac output, C) ejection fraction and D) fractional shortening of both WA and CA zebrafish at two acute temperatures. Error bars represent SEM. * represents significant difference of both groups between temperatures. † represents significant interacting effect of acclimation state. Red: warm acclimated (WA) zebrafish. Blue: cold acclimated (CA) zebrafish.

Fractional shortening (FS), the degree of shortening of the ventricle between end-diastole and end-systole, is another index of myocardial contractility. Acute temperature change did not change FS significantly, but the acclimation state did have a confounding effect on the temperature response (cross-effect p=0.006). At 28ºC, both CA and WA had similar FS (≈17%), but at 18ºC CA fish had a significant higher FS (CA = 23% vs WA = 16% - Figure 4-4d).
In the WA group neither end systolic volume (ESV) nor EDV changed significantly with acute temperature fluctuation (ESV of 1.24 µl and 1.19 µl, EDV of 2.34 µl and 2.31 µl at 18°C and 28°C, respectively). With CA, both ESV and EDV show a lower trend (ESV of 0.99 µl and 1.05 µl, EDV of 2.16 µl and 2.11 µl at 18°C and 28°C, respectively) but there was no significant effect of acclimation state on the acute temperature fluctuation response seen in either parameter. It is important to note that SV in some studies is normalized to body weight and denoted as µl/g. This is a common normalization procedure seen in many teleost studies (goldfish (17); flounder (30)) due to the correlation between blood volume and body weight. However, since no significant changes in body weight were seen between cold-acclimated and warm-acclimated fish used in the echocardiography study the data are shown in absolute values. For comparison with other studies and species, normalized CO is displayed as ml·min⁻¹·kg⁻¹ in Figure 4-4.

### Table 4-1 Parameters from echocardiographic B-mode of the zebrafish heart for WA vs CA fish at 18°C and 28°C

<table>
<thead>
<tr>
<th>Group</th>
<th>Temp °C</th>
<th>HR (bpm)</th>
<th>BW (g)</th>
<th>SV (µl)</th>
<th>EF (%)</th>
<th>FS (%)</th>
<th>ESV (µl)</th>
<th>EDV (µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>SEM</td>
<td>mean</td>
<td>mean</td>
<td>mean</td>
<td>mean</td>
<td>mean</td>
<td>mean</td>
</tr>
<tr>
<td>WA</td>
<td>18</td>
<td>78</td>
<td>6</td>
<td>0.6</td>
<td>1.1</td>
<td>0.1</td>
<td>46.9</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>162</td>
<td>10</td>
<td>0.6</td>
<td>1.1</td>
<td>0.1</td>
<td>48.1</td>
<td>0.8</td>
</tr>
<tr>
<td>CA</td>
<td>18</td>
<td>80</td>
<td>8</td>
<td>0.6</td>
<td>1.2</td>
<td>0.1</td>
<td>54.2</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>169</td>
<td>15</td>
<td>0.6</td>
<td>1.1</td>
<td>0.1</td>
<td>49.9</td>
<td>0.9</td>
</tr>
</tbody>
</table>

HR = heart rate in bpm. BW = body weight in g. SV = stroke volume in µL. EF = ejection fraction (%). FS = fractional shortening (%). ESV = end systolic volume in µL. EDV = end diastolic volume in µL.

* represents significant difference between temperatures within group p<0.05  † represents significant interacting effect of the acclimation state p<0.05

**4.4.2. Pulse wave Doppler measurements**

Ventricular inflow velocities including E and A (early and atrial) peak velocity and ventricular outflow velocities were significantly greater at 28°C (p<0.0001 - Table 2) in both groups. There were no significant differences in A peak or V outflow velocities between WA and CA fish when measured at either 18°C or 28°C. There was a confounding effect of acclimation state on E peak velocity temperature response, shown as WA fish at 28°C having a greater E peak velocity than CA fish (WA 52.34 mm/s vs CA 32.02 mm/s, p =
0.004, Figure 4-5a). Acute temperature decrease significantly lengthens the duration of both the E phase ($p = 0.0076$) and A phase ($p<0.0001$). The change in duration of only the A phase is confounded by acclimation state (Table 4-2a). Both E duration and A duration are part of the overall increase in inflow time with acute temperature decrease ($p = 0.0007$). When normalized to beat duration (HR$^{-1}$), the acute temperature no longer significantly affected duration time (Table 4-2b).

4.4.3. E/A values

No significant variation in the E/A ratio was found with acute temperature fluctuation. E/A ratios were comparable at either temperature in WA fish ($EA \approx 0.28$). While the effects of acclimation did not show statistical interaction with those of acute temperature response, a trend of lower E/A ratios appeared in CA fish (Figure 4-5c).

4.4.4. The Myocardial Performance Index (MPI)

The myocardial performance index (MPI) is a pulse wave Doppler-derived time interval index that incorporates both systolic and diastolic time intervals and is impacted by both systolic and diastolic cardiac function. Both systolic and diastolic dysfunction can result in abnormality in myocardial relaxation which can prolong the relaxation period (isovolumic relaxation time, IVRT). Systolic performance, reflected in part by the isovolumic contraction time (IVCT) was longer at 18 vs. 28ºC for both WA and CA fish ($p<0.0001$). Acclimation state did affect the acute temperature response (cross-effect $p = 0.019$). IVCT was longer in WA fish relative to CA at both 18ºC and 28ºC (Figure 4-5d). Diastolic performance, reflected in part by the isovolumic relaxation time was also longer at 18 vs. 28ºC, a significant effect from the acclimation condition ($p<0.0001$). However, while IVRT appeared to be longer in CA fish, this trend was not statistically significant. Ejection time (ET) was also significantly decreased with an acute temperature decrease to 18ºC ($p<0.0001$) but was unaffected by acclimation state. Overall, the MPI did not vary at either acute temperature condition, and there was no acclimation effect (Table 4-2c).
Table 4-2  Parameters from pulse wave Doppler and colour Doppler imaging of ZF heart

A. 

<table>
<thead>
<tr>
<th>Group</th>
<th>Temp (°C)</th>
<th>E peak velocity (mm/s) mean</th>
<th>SEM</th>
<th>A peak velocity (mm/s) mean</th>
<th>SEM</th>
<th>Outflow peak velocity (mm/s) mean</th>
<th>SEM</th>
<th>E flow duration (ms) mean</th>
<th>SEM</th>
<th>A flow duration (ms) mean</th>
<th>SEM</th>
<th>Inflow duration (ms) mean</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>WA</td>
<td>18</td>
<td>30.5</td>
<td>2.1</td>
<td>115.6</td>
<td>12.3</td>
<td>126.5</td>
<td>17.1</td>
<td>290.3</td>
<td>68.2</td>
<td>118.2</td>
<td>3.3</td>
<td>408.5</td>
<td>68.7</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>52.4</td>
<td>4.7</td>
<td>212.7</td>
<td>23.9</td>
<td>256.4</td>
<td>32.9</td>
<td>139.7</td>
<td>16.3</td>
<td>58.9</td>
<td>1.0</td>
<td>198.7</td>
<td>16.6</td>
</tr>
<tr>
<td>CA</td>
<td>18</td>
<td>27.7</td>
<td>2.3</td>
<td>129.5</td>
<td>11.5</td>
<td>128.7</td>
<td>11.2</td>
<td>354.1</td>
<td>87.6</td>
<td>102.8</td>
<td>1.4</td>
<td>456.9</td>
<td>86.6</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>32.0</td>
<td>3.6</td>
<td>219.6</td>
<td>26.3</td>
<td>193.5</td>
<td>10.9</td>
<td>162.9</td>
<td>44.9</td>
<td>57.3</td>
<td>2.1</td>
<td>220.2</td>
<td>44.9</td>
</tr>
</tbody>
</table>

B. 

<table>
<thead>
<tr>
<th>Group</th>
<th>Temp (°C)</th>
<th>Beat duration (ms) mean</th>
<th>SEM</th>
<th>E flow duration (ms) mean</th>
<th>SEM</th>
<th>A flow duration (ms) mean</th>
<th>SEM</th>
<th>Inflow duration (ms) mean</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>WA</td>
<td>18</td>
<td>814.0</td>
<td>74.3</td>
<td>32.1</td>
<td>4.5</td>
<td>15.4</td>
<td>1.2</td>
<td>47.5</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>381.7</td>
<td>22.0</td>
<td>35.7</td>
<td>2.4</td>
<td>15.9</td>
<td>0.8</td>
<td>51.5</td>
<td>1.9</td>
</tr>
<tr>
<td>CA</td>
<td>18</td>
<td>834.2</td>
<td>88.5</td>
<td>36.9</td>
<td>6.7</td>
<td>13.7</td>
<td>1.5</td>
<td>50.6</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>387.2</td>
<td>42.6</td>
<td>36.5</td>
<td>6.0</td>
<td>16.1</td>
<td>1.3</td>
<td>52.6</td>
<td>4.7</td>
</tr>
</tbody>
</table>

C. 

<table>
<thead>
<tr>
<th>Group</th>
<th>Temp (°C)</th>
<th>IVCT (ms) mean</th>
<th>SEM</th>
<th>IVRT (ms) mean</th>
<th>SEM</th>
<th>ET (ms) mean</th>
<th>SEM</th>
<th>MPI (%) mean</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>WA</td>
<td>18</td>
<td>50.1</td>
<td>2.8</td>
<td>110.4</td>
<td>4.2</td>
<td>236.0</td>
<td>9.3</td>
<td>0.69</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>28.5</td>
<td>1.2</td>
<td>39.3</td>
<td>1.8</td>
<td>116.5</td>
<td>7.0</td>
<td>0.60</td>
<td>0.04</td>
</tr>
<tr>
<td>CA</td>
<td>18</td>
<td>36.2</td>
<td>3.2</td>
<td>116.7</td>
<td>4.3</td>
<td>246.7</td>
<td>11.7</td>
<td>0.62</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>27.2</td>
<td>2.6</td>
<td>42.5</td>
<td>1.6</td>
<td>111.2</td>
<td>5.1</td>
<td>0.64</td>
<td>0.04</td>
</tr>
</tbody>
</table>

A. Ventricular inflow, made up of two components: early (E) filling peak velocity occurring during ventricle relaxation (E peak vel in mm/s) and the atrial (A) filling peak velocity resulting from active atrial contraction (A peak vel in mm/s), and ventricular outflow peak velocity (V outflow vel in mm/s). Early flow duration (E duration in ms), late flow duration (A duration in ms) make up the total inflow duration. B. Normalized duration times. Duration is expressed as proportion of beat duration (ms) calculation as the inverse of HR. C. Isovolumic contraction time measuring systolic performance (IVCT) in ms, isovolumic relaxation time measuring diastolic performance (IVRT) in ms and ejection time (ET) in ms are also measured from this view. The myocardial performance index (MPI) is a Doppler-derived time interval index that combines both systolic and diastolic cardiac performance (IVCT + IVRT)/ET.

* represents significant difference between temperatures p<0.05
† represents significant interacting effect of acclimation state p<0.05
Figure 4-5  Ventricular inflow, made up of two components: early (E) filling peak velocity occurring during ventricle relaxation (E peak vel in mm/s) and the atrial (A) filling peak velocity resulting from active atrial contraction (A peak vel in mm/s), and ventricular outflow peak velocity (V outflow vel in mm/s).

Early flow duration (E duration in ms), late flow duration (A duration in ms) make up the total inflow duration. B. Normalized duration times. Duration is expressed as proportion of beat duration (ms) calculation as the inverse of HR. C. Isovolumic contraction time measuring systolic performance (IVCT) in ms, isovolumic relaxation time measuring diastolic performance (IVRT) in ms and ejection time (ET) in ms are also measured from this view. The myocardial performance index (MPI) is a Doppler-derived time interval index that combines both systolic and diastolic cardiac performance (IVCT + IVRT)/ET.

* represents significant difference between temperatures p<0.05.
† represents significant interacting effect of acclimation state p<0.05.
4.5. Discussion

In this study high frequency echocardiographic measurements allowed us to obtain a critical assessment of in vivo cardiac function in zebrafish as a function of temperature. The zebrafish has been used in the literature as a robust model of vertebrate cardiac development and function because of the availability of annotated genomic information, accessibility of numerous mutant strains and the transparency of the developing embryos that permits measurement of numerous cardiac parameters by visual inspection. Due to loss of transparency in adulthood as well as the small size of adult zebrafish hearts (1-2 mm in diameter), it is challenging to measure all facets of cardiac function in vivo without sophisticated imaging techniques. Detailed, reproducible measurements using high-frequency echocardiography in zebrafish can be taken only with the use of a very high frequency probe (70 MHz) and extreme diligence. In the literature, variability exists in in vivo fish cardiac function results, likely due to the method of anesthetization and restraint of fish in addition to lower resolution echocardiographic measurements. The anesthetic protocol used in this study, modified from Huang et al. (23), allows for prolonged state of anesthesia with a significantly lower impact on both heart rate (HR) and ejection fraction (EF) (23). All fish maintained a minimal EF of 40\% throughout the experiment. While at first glance the HR may seem to vary considerably especially with the acute temperature decrease, all temperature-dependent changes in HR correspond to a $Q_{10}$ in the range of 2.2. Utilizing $Q_{10}$ to express acute temperature changes,

$$Q_{10} = \frac{R_2}{R_1} \exp \frac{10}{T_2-T_1}$$

or simply the fold-change in activity per 10°C, the data fell in the range of 2.2. These $Q_{10}$ values are well within the expected normalized rate change of two-to three-fold for physiological processes with increasing temperature. The echocardiographic measurements in this study show a similar range in HR in vivo, as well as a comparable change in HR with cooling (32). Overall, these inclusionary criteria suggest cardiac function was maintained to ideal conditions to monitor changes in response to a stressor such as temperature change.
4.5.1. Heart rate is primary cardiac function response to acute temperature changes

Resting HR in ectotherms is greatly influenced by temperature (10, 13). Typically, teleosts cope with acute temperature changes by modifying HR to maintain CO (zebrafish (33); cod – (19); salmon (6); flounder –(34)). Most temperate fish exhibit resting heart rate increases with increased temperatures on the average with a Q10 of 2.0 (47). In this study, an acute decrease of temperature from 28°C to 18°C resulted in a 48% decrease in HR, demonstrating a Q10 of 2.2. Stroke volume, the other determinant of CO, undergoes very little change in response to acute temperature perturbations with a Q10 of ~1.0 (6, 19, 34, 43). Modulations in HR alone appear to be responsible for the changes in cardiac output reported in this study (Q10 = 2.2). These results corroborate what has been suggested from studies with other teleost species (33) using in situ and other approaches. EDV was unchanged by cooling from 28°C to 18°C, similar to what has been reported with variable temperatures in measurements from other teleosts (2, 16) using other methodologies. Increased EDV would have a particularly important role in determining CO in fish, as it may reduce the scope for increasing SV (2, 8, 16, 29). The lack in change of both SV and EDV with cooling supports the notion that the cardiac output of zebrafish is controlled, to a great extent, by HR in response to acute temperature fluctuations.

Ventricular inflow and outflow velocities were both significantly increased at 28°C relative to 18°C (Q10 – 1.8 and 2.0, respectively; Table 4-2). Our data showed a slightly higher ventricular outflow than ventricular inflow at 28°C (213 mm·sec⁻¹ inflow vs 256 mm·sec⁻¹ outflow) and similar velocities for both inflow and outflow at 18°C (115.6 mm/sec inflow and 126.5 mm·sec⁻¹ outflow). Sun et al (46) found peak ventricular outflow (800 mm·sec⁻¹) to be significantly less than peak ventricular inflow (1440 mm·sec⁻¹), demonstrating the opposite of what was seen in this study. In both the Sun et al. and Hu et al. (21, 46) studies, velocity measurements were determined at a lower temperature range (15-25°C) with high to very high doses (100 to >200 ppm) of MS-222 as an anesthetic. Any alteration in flow duration with acute temperature change corresponded with fluctuations in beat duration, or inversely proportional to HR (Table 4-2b), both of which would be affected by MS-222. As well, our study optimized the locations to measure velocity (the point between atrium and ventricle for inflow velocity and between ventricle
and bulbus for outflow velocity) and used color Doppler for visualization of the direction of blood flow to improve the resolution of measurement. This technique of combining colour Doppler with pulse-wave Doppler is more user-friendly than using only pulse-waved Doppler where estimation of an average angle is required for each zebrafish (26, 30). The results presented here demonstrate Q_{10} values for chamber-to-chamber velocities are internally consistent with variations in other parameters measured (HR, CO).

4.5.2. Zebrafish CO shows no response to cold acclimation

In the current study acclimation had no confounding effect on CO, either through HR or SV. Both CA and WA fish had comparable HR at 18^\circ C and 28^\circ C, suggesting that the HR response to the current ambient temperature outweighs any acclimated response. SV is highest when the ambient temperature matches the acclimation temperature in this study, such as SV in the CA group is higher at 18^\circ C while SV in the WA group decreases upon cooling to 18^\circ C. The variation between the two resulted in a slightly greater Q_{10} for CO in WA fish compared to CA fish (2.15 vs 1.89). This suggests that there is a slightly greater compensation over the 10^\circ C acute change is made by WA fish. While SV may be altered with cold acclimation of sufficient duration and temperature change in many teleost species (3, 5, 28, 39), this does not appear to be an acclimation response here. These results correspond to overall lack of ventricular hypertrophy in zebrafish measured histologically in other studies (25), in contrast with larger hearts of cold-acclimated trout exhibiting enhanced SV (9, 27). The possibility of species-specific responses to acclimation makes characterization of functional properties of the heart in adult zebrafish crucial as its role as a model for cardiac function expands.

While changes in HR with acute temperature change are reflected in both IVCT and IVRT, these measures did display a confounding effect of acclimation state (Table 4-2). Relaxation time calculated as IVRT is longer with an acute temperature decrease and tends to be lengthened in cold-acclimated fish though not to a significant extent. Predictably, in acute temperature fluctuation conditions, contraction time calculated as IVCT is longer at 18^\circ C. This acute temperature response is confounded by acclimation state, with CA fish exhibiting shorter IVCT at both temperatures. The biggest difference is
that CA fish demonstrated a much shorter IVCT than WA fish at warm temperatures. This could be indicative of a greater degree of difficulty for ectotherms to deal with temperature increase compared to temperature decrease (11, 31, 40).

In cardiac muscle, cold-acclimation can extensively modify expressions of genes involved with protein synthesis and hypertrophy as well as genes encoding for contractile proteins (1, 15, 18, 25). In zebrafish specifically, the acclimation protocol used in this study increases the mRNA levels of cardiac troponin C (cTnC) with decreased temperature, a critical component of the cardiac contractile element (18). Conditional antisense knockdown of cTnC in zebrafish results in a significantly lower EF (37), a common index of contractility (32). This study revealed a significant increase in EF with cold acclimation, which is consistent with an increase in the expression of contractile proteins such as cTnC (18). Importantly, changes to EF were only observed at cold temperatures, where CA fish at 18ºC had significantly higher EF than CA fish at 28ºC or WA fish at either temperature. This suggests that tissue remodeling optimized the contractility in CA fish to function at 18ºC even if overt morphological changes were not seen. Using histological techniques, other zebrafish studies have shown no change in heart volume but a decrease in the compact myocardium with cold acclimation (25). Increases in compact myocardium is suspected to be compensation for decreased power-generating ability or simply increased activity levels at higher temperatures (15), and may also account for a lack in change of size of the ventricle as a whole. Further work will need to be done to link potential changes in contractility with the molecular mechanisms of remodelling specific to zebrafish.

4.5.3. Filling patterns do not change in response to acute temperature or acclimation

In this study, the E/A ratio was found to be 0.28 regardless of the temperature. This value corroborates previous zebrafish studies, which have shown similar E/A ratios at 15ºC (20, 46) and 25ºC. The E/A ratio accounts for the two components to the preload of the ventricle, early filling created by E-flow and late filling created by A-flow which is atrial-contraction dependent (Figure 4-5). These E/A ratios indicate that late diastolic filling caused by atrial contraction in zebrafish account for the majority of ventricular filling
regardless of the temperature. In mammals, this phase is created by venous pressure and under resting conditions is the main determinant of ventricular filling. Ventricular inflow in fish, unlike that in mammals, is primarily due to atrial contraction (A phase) (12), seen as a larger A-phase peak velocity compared to the E-flow. This model of filling is supported by the similarity between atrial and ventricular volume (12, 46), and the favorable filling pressure gradient that exists only during atrial systole (45). Both E–flow and A-flow significantly decrease as temperature changes from 28ºC to 18ºC with comparable Q_10 values (A peak velocity 1.84; E peak velocity 1.71). Acute reduction in temperature has been shown to have larger effects in atrial versus ventricular action potential duration (32), but this does not appear to cause a difference in the proportion of active atrial contraction that contributes to EDV. Since both E and A flow change with temperature, the E/A ratio is an important parameter in evaluating cardiac function for comparison across species and between studies.

Acclimation state did have statistical effect on E/A fraction with decreased ratios seen with cold acclimation. This E/A ratio suggests that cold-acclimated fish rely more on active atrial contraction to maintain SV. There was a significant cross-effect of acclimation on E-peak velocity response to acute temperature. CA fish had lower E-flow. The Q_{10} for the change in E-flow velocity in WA fish was 1.71, but in CA fish was only 1.15. Lack of compensation in passive filling patterns with acclimation is suggested by this ratio and could also be reflective of a less compliant ventricle suggested by EDV and IVRT in CA. Compliance of the zebrafish ventricle has been suggested to lower than that of human hearts (25, 26) this is turn is influenced by temperature. Our current results support that the passive properties of the ventricle may be influenced by acclimation even with no change in ventricular size, such as is seen in remodelling of myocardial layers seen with acclimation in other zebrafish studies (25). An increase in temperature is more stressful for an ectothermic fish to maintain cardiac output (11) which may be responsible for CA fish facing acute temperature increase to slightly modulate E/A ratio. In fact, CA fish at 28ºC demonstrate the lowest E/A ratio, possibly indicating ventricular dysfunction. With a larger EF in CA as well as a slightly larger EDV, there is a possibility of a greater atrial contraction or simply a longer cardiac cycle suggested by IVRT.
4.6. Summary

This study allowed for the critical assessment of cardiac parameters in the intact zebrafish in response to both acute and long-term temperature change using high-resolution echocardiography. Modulations in HR dictate changes in cardiac output in response to acute temperature changes in both WA and CA fish. SV was highest when the ambient temperature matches the acclimation temperature. All other functional parameters were maintained throughout the acute temperature challenge. Cold acclimation improved systolic function at 18°C in comparison to the WA group with increases in both EF and FS and decreases in IVCT. The decreased E velocity and E/A ratio in the CA group may be associated with increased reliance on atrial contraction. However, the consistency of the E/A ratio with respect to temperature shows it is an important parameter in evaluating cardiac function for comparison between studies in zebrafish. The accuracy of the in-vivo high-resolution echocardiography used in this study lends confidence to our understanding of responses of teleost cardiac function to environmental perturbation. Species-specific responses to both acute temperature changes and longer-term acclimation make this high-resolution characterization of functional properties of the heart critical for the expansion of the adult zebrafish as a model for cardiac function.

4.7. Acknowledgements

These studies were supported by an NSERC Discovery grant to GFT and a Collaborative Health Research Program grant to GFT and MVS. MVS and MFB are Michael Smith Foundation for Health Research Scholars and GFT is a Tier 1 Canada Research Chair. The assistance of Bruce Leighton for set up of the fish acclimation system as well as Haruyo Kashihara for assistance with fish care and transport is greatly appreciated.
4.8. References


Chapter 5.
Sub-functionalization of two distinct troponin C paralogs of the teleost heart in a chamber and temperature dependent manner

This chapter is largely based on the following original research article:


This study was fully conceived and written by me under the advisement of the co-authors.
5.1. Abstract

The teleost-specific whole genome duplication created multiple copies of genes allowing for sub-functionalization of isoforms. In this study, we show that the teleost cardiac Ca$^{2+}$-binding troponin C (TnC) is the product of two distinct genes: cardiac TnC (cTnC, \textit{TnnC1a}) and a fish-specific slow skeletal TnC (ssTnC, \textit{TnnC1b}). The TnC1b gene is novel to teleosts as mammals have a single gene commonly referred as cTnC but which is also expressed in slow skeletal muscle. In teleosts, the data strongly indicate that these are two TnC genes are different paralogs. Because we determined that TnC1b exists across many teleosts but not in basal ray-finned fish (e.g. bichir), we propose that these paralogs are the result of an ancestral tandem gene duplication persisting only in teleosts. Quantification of mRNA levels was used to demonstrate distinct expression localization patterns of the paralogs within the chambers of the heart. In the adult zebrafish acclimated at 28°C, TnC1b mRNA levels are two-fold greater than TnC1a mRNA levels in the atrium, whereas TnC1a mRNA was almost exclusively expressed in the ventricle. Meanwhile, rainbow trout acclimated at 5°C showed TnC1a mRNA levels in both chambers significantly greater than TnC1b. Distinct responses to temperature acclimation were also quantified in both adult zebrafish and rainbow trout, with mRNA in both chambers shifting to express higher levels of TnC1a in 18°C acclimated zebrafish and 5°C acclimated trout. Regulatory sub-functionalization of TnC isoforms may provide insight into how teleosts achieve physiological versatility in chamber-specific contractile properties.

5.2. Introduction

Teleost fish exhibit a wide variation in thermal tolerance and a capacity to acclimate to changing temperatures in order to cope with both longer term seasonal and acute temperature changes (42). The fish heart shows variable tolerance and plasticity to temperature (13) but there are many species-specific differences in these responses possibly correlating with habitat conditions and thermal preferences (38). The breadth of interspecies variation within teleost species is in stark contrast with cardiac muscle of
endothemic mammalian species, making the temperature-induced adaptations of the contractile unit of teleost cardiac muscle intriguing.

The teleost heart is composed of four chambers arranged in series: venous sinus, atrium, ventricle and bulbus arteriosus. The two contractile chambers acting as pumps are the atrium and ventricle, a simplified version of that seen in tetrapods. The atrium and the ventricle differ not only morphologically but also physiologically with distinct contractility characteristics. The end-diastolic volume (EDV) in fish is largely determined by atrial contraction rather than central venous pressure as seen in mammals (9). This overall structure allows for increased reliance on changes in stroke volume (SV) to respond to greater hemodynamic loads. Because of this increased involvement of the atrium in driving ventricular filling and hence cardiac output, the fish-specific contractile properties of the atria and the ventricles are likely to be important in achieving a higher degree of physiological versatility (16).

Temperature is a striking environmental condition that greatly impacts cardiac contractility in ectothermic species. Factors such as heart rate, cardiac output have all been shown to fluctuate in species such as salmonids with changing temperature. With decreased temperature, the Ca\(^{2+}\) sensitivity of the cardiac contractile unit decreases in both mammals (23) and fish (8) despite myofilament Ca\(^{2+}\) sensitivity being a critical factor in regulating cardiac contractility. It has been postulated that ectotherms require heightened myofilament Ca\(^{2+}\) sensitivity to maintain cardiac function through changing environmental conditions. Skinned salmonid cardiomyocytes have been shown to have a significantly higher Ca\(^{2+}\) sensitivity at any given temperature (over the measured range of 2\(^{\circ}\)C to 28\(^{\circ}\)C) than those from mammals (8). A key regulator of myofilament Ca\(^{2+}\) sensitivity, the cardiac troponin (cTn) complex, is made up of three proteins (cTnC, cTnI, cTnT). Troponin C, the Ca\(^{2+}\) binding protein that initiates myocyte contraction, has been shown to play a role in enhancing myofilament Ca\(^{2+}\) sensitivity in fish (18). TnC consists of two globular domains connected by an α-helical linker with each domain having two EF-hand Ca\(^{2+}\) binding sites. In mammalian fast skeletal TnC (fsTnC), all four EF hand sites bind Ca\(^{2+}\) with affinities that are physiologically relevant. The C-terminal sites are considered high-affinity Ca\(^{2+}/Mg\(^{2+}\) sites and are known to be structurally integral to anchoring TnC to the troponin complex, while the N-terminal domain sites are relatively low-affinity but
Ca$^{2+}$-specific. Cardiac TnC possesses only one low affinity Ca$^{2+}$ binding site (site II) as site I has been rendered non-functional by amino acid substitutions which preclude Ca$^{2+}$ coordination. In cardiac muscle Ca$^{2+}$ binding to site II initiates a complex change in the interrelations of the troponin molecules on the thin filament initiating force production of the contractile element, making cTnC a key regulator of myofilament Ca$^{2+}$ sensitivity (35).

The importance of the role of cTnC is demonstrated by its high degree of sequence conservation across phylogenetically diverse groups of organisms and millions of years of evolution. In fact, across mammals, cTnC amino acid sequences are almost completely identical while mammalian sequences relative to that of trout show 92% amino acid identity (20). Because of the key role of this protein in cardiac physiology, modifications in the sequence of amino acids resulting in functional divergence of troponin C can alter the entire contractile reaction of the myocyte. Despite the highly conserved nature of TnC, subtle variations are seen as interspecies orthologs and intraspecies paralogs between tissues. In trout cTnC, specific amino acids have been shown to be responsible for the increased Ca$^{2+}$ affinity seen relative to mammals, including Asn2, Ile28, Gln29, and Asp30 (NIQD). These residues are hypothesized to allosterically affect the ability of site II to bind Ca$^{2+}$ (22). This increased Ca$^{2+}$ sensitivity is observed despite the fact that fish cTnC are temperature-sensitive. These amino acid substitutions enable greater Ca$^{2+}$ binding at lower temperatures than would otherwise be possible (21).

However, with the teleost-specific whole genome duplication (29), the possibility exists that multiple isoforms of genes exist that may be expressed differentially on a temporal or spatial scale or in response to environmental factors such as temperature. While most genes are lost after a duplication event, tissue specific expression may lead to increased retention of multiple isoforms since paralogs can divide up ancestral function (sub-functionalization) or assume new functions (neo-functionalization) (14). Sub-functionalization can manifest either in a regulatory or structural nature and in turn can lead to further neo-functionalization (36). In zebrafish embryos two separate paralogs have been identified as being comparable to mammalian cTnC: one primarily expressed in the heart (TnnC1a or cTnC) and the other in slow skeletal muscle (TnnC1b or ssTnC)(40). One important note on nomenclature is that TnnC1a and TnnC1b are
consensus terms used by the Zebrafish Community's information network (zfin.org) while the terms cTnC and ssTnC are consensus terms in the mammalian literature. These genes are both located on chromosome 23 in zebrafish (TnC1a: 4055544-4060124; TnC1b: 19608998-19611413) but display independent gene expression patterns both temporally and spatially between tissues. This expression localization pattern suggests differences in functionality between the two paralogs. However, thus far this pattern has only been seen in embryo with no information on adult zebrafish paralog distribution, or in fact the presence of any orthologs in any other teleost species.

Through phylogenetic analysis of existing fish genome sequences, we demonstrate that TnC1b is not only present in zebrafish (as seen in embryos) but broadly across teleost species. Using adult zebrafish and rainbow trout as models, we examined paralog gene expression tissue localization patterns as well as temperature acclimation effects to begin to examine the relevance of three fish-specific paralogs of TnC. Our focus was on the two paralogs expressed in cardiac tissue and their relation to atrio-ventricular differences in contractility (16). We show that the TnC1b and TnC1a transcripts are expressed in a chamber-specific manner in both the zebrafish and rainbow trout hearts, indicating a possible involvement in functional differences in the atrium and ventricle like other chamber specific isoforms of contractile proteins (e.g. myosin heavy chain and myosin light chain; (28, 45)). As well, we found both chambers have a unique response of relative paralog expression to an environmental perturbation: temperature acclimation.

5.3. Methods

5.3.1. Sequence and phylogenetic analyses

All available TnC gene sequences homologous to zebrafish cTnC or ssTnC were found using the National Center for Biotechnology Information (Bethesda, MD) non-redundant protein database (http://www.ncbi.nlm.nih.gov/guide/proteins/), the Ensembl Genome Browser (Welcome Trust Genome Campus, Hinxton, Cambridge, UK) (http://uswest.ensembl.org/index.html), and GenBank (http://www.ncbi.nlm.nih.gov/genbank/).
BLAST searches (2) were set up for several species EST databases against zebrafish cTnC (TnnC1a) and ssTnC (TnnC1b) sequences (GI:28822162 and GI:50344823, respectively).

Multiple amino acid sequence alignments were performed using MUSCLE (Multiple Sequence Comparison by Log-Expectation; (12)) via MEGA5 (Molecular Evolutionary Genetics Analysis 5.0(43)). The evolutionary histories for these amino acid sequences were inferred using the maximum likelihood algorithm (MEGA5) and Bayesian inference (Topali v2.5). Analyses were based on the JTT-matrix model (27), the optimal model of evolution built on the Akaike Information Criterion (25). For maximum likelihood, the bootstrap consensus trees were inferred from 500 replicates. For Bayesian inference, analysis was run for 1 x 106 generations with Metropolis-coupled MCMC using markov chains generation sampled every 100 generations. A conservative 25% of the trees were discarded as burn-in for Bayesian analysis. All trees were drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 161 residues in the final dataset.

5.3.2. Animals and cold acclimation

All animal experiments were approved by the Simon Fraser University Animal Care Committee (protocol number 989K-90). Adult zebrafish purchased from a local supplier were maintained at a 12-h:12-h light:dark photoperiod at 28°C in dechlorinated water in a flow-through tank and fed ad libitum (Nutrafin Max, Hagen Inc, Baie d’Urfé, QC, Canada). Rainbow trout were held in an outdoor facility in dechlorinated water and fed ad libitum. For the thermal acclimation group of adult zebrafish, after being held at 28°C for two weeks in flow-through tanks, the temperature was decreased by 2.5°C/week until the end-point temperature of 18°C was reached. Fish were then held at this temperature for 3 weeks. Trout were allowed to seasonally shift with ambient temperature and light cycles, but cardiac and skeletal muscle data were collected at a temperature of 5°C during the winter months and 15°C during the summer.
Zebrafish were sacrificed via cold shock followed by decapitation. Trout were sacrificed by a swift blow to the skull followed by severing the spinal cord. Cardiac tissue from whole heart and skeletal tissue from the lateral line was taken from a first group of fish (28°C in zebrafish and 5°C in trout). For the remaining warm and cold-acclimated fish, the heart was removed via a small triangular section that was excised from the zebrafish along the gill arch to the head. From this section the whole heart was teased out by grabbing onto the bulbus arteriosus under a dissecting scope. After the bulbus arteriosus was removed, the atrium and the ventricle were physically separated using scissors (39). In the trout, the larger size allowed for more traditional dissection through the ventral side. All samples were placed immediately into cold Trizol (Invitrogen, Burlington, ON). Seven fish were pooled for each individual sample and then each sample was frozen at -80°C and stored at this temperature until use.

5.3.3. RNA analysis

Total RNA was extracted from each sample using Trizol followed by RNeasy kit (Qiagen, Mississauga, ON) according to the manufacturer’s instructions. Samples were quantified spectrophotometrically at 260 nm before storage at −80°C. RNA purity was assessed by measuring absorbance at 280 nm to assess protein contamination, with any samples having the 260/280 ratio falling outside the range of 1.9-2.1 being a criterion for being discarded. Genomic DNA removal and reverse transcription of RNA samples were done using the Qiagen Quantitect Reverse Transcription kit according to the manufacturer’s instructions with each reaction scaled to 0.5 μg RNA.

Real time quantitative PCR (qRT-PCR) analysis was performed on a BioRad CFX96 Touch Real-Time PCR System (BioRad, Mississauga, ON) using the following conditions: an initial denaturation for 10 min at 95°C followed by 40 cycles of 15 s denaturation at 95°C, 30 s annealing at optimal primer temperature (Table 5-1) and 36 s extension at 72°C. Samples were assayed in duplicate in a 20 μl reaction volume containing 10 ng cDNA, 12 μl SsoAdvanced SYBR Green Supermix (BioRad) and 0.25 μmol of each primer. Negative controls (no template or selected untranscribed RNA) were
run as well to ensure the absence of contamination. Select control samples were re-assayed to ensure no significant difference between assays.

Analysis was performed according to the \( \Delta\Delta C_t \) method using the geometric mean of two distinct housekeeping genes to normalize the data (44). β-actin was originally chosen as an appropriate reference gene for zebrafish (6) and confirmed by both dissociation curves and the stability of Ct values across conditions and samples. The second gene, elongation factor-1 alpha, was measured using two primer sets, one spanning exons 3 and 4, and the second spanning exons 5 and 6, allowing the comparison of values from each primer set to check for RNA stability (26)- second gene set data not shown). Specific primers for each gene were designed using a combination of Primer3 (37) and Oligonucleotide Properties Calculator (30) to amplify a single product (using zebrafish and rainbow trout published sequences - see Table 5-1), as checked by regular PCR and dissociation curve analysis post-real-time PCR. Products were confirmed via sequencing (CMMT / CFRI DNA Sequencing Core Facility, Vancouver, BC). All data are presented as mean values ± s.e.m., and are corrected for by using the geometric mean of β-actin and ef1a 1, then expressed relative to the isoform with either lower expression (Figure 5-4) or TnC1b (Figure 5-5). Significance was determined between chambers by p values less than 0.05 by unpaired Student's t-tests.
Table 5-1  Primers used for qRT-PCR analyses of zebrafish (ZF) and rainbow trout (RT).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Accession Number</th>
<th>Primer F</th>
<th>Primer R</th>
<th>Product Size</th>
<th>Tm</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZF TnC1a</td>
<td>AF434188</td>
<td>AAGCAGGCGGAGAGAAGAAGCTCTTCAGGTTGTTAC</td>
<td>GCTCTTCTTTGAGGATGAGTTC</td>
<td>153</td>
<td>59°C</td>
</tr>
<tr>
<td>ZF TnC1b</td>
<td>NM_001002085</td>
<td>AGCAGGCTTGGAGAACTTG</td>
<td>GCTCTTCTTTGAGGATGAGTTTT</td>
<td>152</td>
<td>59°C</td>
</tr>
<tr>
<td>ZF β-actin</td>
<td>AF057040</td>
<td>TTCTGGGTCTACTGTGTTGT</td>
<td>ATCTCTACAGGTAGTCGGGTCA</td>
<td>170</td>
<td>59°C</td>
</tr>
<tr>
<td>ZF ef1a</td>
<td></td>
<td>TGAAATTCGAGACGAGC</td>
<td>GGTCTGTCCCTTCTGAGA</td>
<td>168</td>
<td>59°C</td>
</tr>
<tr>
<td>RT TnC1a</td>
<td>AY281129</td>
<td>CTCAAAAGCAGCGGTAGAAGC</td>
<td>GATCTGACTGCCATCTCTA</td>
<td>203</td>
<td>59°C</td>
</tr>
<tr>
<td>RT TnC1b</td>
<td>BX863663.3</td>
<td>AAGCAGCAGTGAACCTTTA</td>
<td>GGAACCATCAAGGCTACTGTCC</td>
<td>216</td>
<td>59°C</td>
</tr>
<tr>
<td>RT β-actin</td>
<td>AF157514</td>
<td>TTCAACACCCCCTGCAATGA</td>
<td>CGTCAGGTCCTTCTCCAGGT</td>
<td>204</td>
<td>59°C</td>
</tr>
<tr>
<td>RT ef1a</td>
<td>NM_001124339</td>
<td>CCCCTCCAGGTAGCTACAAA</td>
<td>CACACGCCCAGGGGTAC</td>
<td>210</td>
<td>59°C</td>
</tr>
</tbody>
</table>

5.4. Results

5.4.1. Troponin C sequence variation

High sequence similarity shows troponin C is very well conserved across vertebrates. However, between fish and mammalian sequences there are several variations. While there are 16 amino acid sequence differences (out of a total of 161 residues) between zebrafish TnC1a and mammalian cTnC, there are 19 sequence differences between zebrafish TnC1b and mammalian cTnC (Figure 5-1). There are 18 amino acid sequence substitutions between the zebrafish paralogs TnC1b and TnC1a. There is greater than 93% identity (at the amino acid level) between TnC1a orthologs as well as between TnC1b orthologs, while there is between 87-89% identity between TnC1a and TnC1b paralogs across teleosts (select species shown in Figure 1). Lamprey displays much lower sequence identity than that seen between teleosts (alignment not shown between teleosts and lamprey). The majority of TnC residues critical for Ca\(^{2+}\)-binding (15, 17) are conserved across vertebrates and likewise between TnC1a and TnC1b. Like TnC1a, TnC1b appears to have only site II as a functional Ca\(^{2+}\)-binding domain in the N-
terminal based on the coordinating residues (17). Meanwhile, sites III and IV in the C-terminal domain are conserved across all species in both TnC1a and TnC1b. These sites in all species examined have the X Y Z -Y -X Z sites with residues possessing either a carboxyl or hydroxyl group with the ability to contribute an oxygen atom in order to coordinate Ca\(^{2+}\) (17, 19).

In the N-terminal domain, site II is a region of high sequence conservation due to its importance for the functioning of TnC (20). Site I of all cTnC orthologs displays lower sequence similarity than seen in fsTnC. The insertion of a valine at residue 28 and the replacement of chelating residues Asp29 and Asp31 in all vertebrate cTnCs relative to fsTnC are the reasons for the inability of site I to bind Ca\(^{2+}\) (17). This is seen in Site I of both teleost TnC1a and TnC1b. In both paralogs, certain residues found in teleost hearts such as Gln29 and Asp 30 appear to be teleost-specific traits, but residue 28 shows some variability with Val as seen in mammals and Ile as seen in trout (19). Other residues, such as Asp2 and Glu14, are found in all mammalian cTnC and fish TnC1b but not fish TnC1a. Still others, such as Val4, Asn11, Glu87 and Glu88 appear to be present only in teleost TnC1b.
Figure 5-1  Sequence comparison of representative TnC1a and TnC1b sequences from fish and human.

The Ca²⁺-coordinating positions in each EF-hand site are shown above the sequences (x y z -y-x -z), and the helices are labeled (N, A, C, D, E, F, G, H). Blue residues are fish TnC1 specific, black are human TnC1-like, red are fish TnC1b specific.
5.4.2. Phylogenetic analysis

Due to the length of time since the divergence among vertebrate lineages, the phylogenetic analyses were made based on amino acid sequences rather than nucleotide sequences, in which multiple changes at the third codon positions are likely to have occurred. Tree topologies were consistent with accepted relationships among vertebrates (20) both in maximum likelihood analysis (Figures 5-2 and 5-3) and Bayesian analysis (Figure 5-3). The values reflect the high degree of conservation of TnC with only 18 of 161 residues being variable between isoforms. However, as shown in Figure 2, two distinct TnC clades are seen across vertebrates, with greater variation in the fish lineages in both TnC2 and TnC1. As well as the two traditional groupings across vertebrates, fish TnC1 is split into two branches labeled as TnC1a and TnC1b. For several species of teleost, such as several representative salmonids and the grouper, we have re-annotated them as TnC1b (ssTnC) as opposed to TnC1a (cTnC) as listed in NCBI based upon their grouping and sequence similarity (Figure 3). A more in depth examination of teleost (jawed, bony fish) TnC sequences shows that both paralogs have diverged from the TnC in the jawless lamprey, which is ancestral to teleosts and tetrapods. The basal positions of both the condrichthyes (e.g. catshark) and the ray-finned fish (e.g. bichir) TnC sequence are also consistent with the ancestral state.
Figure 5-2  Phylogenetic tree generated by the maximum likelihood comparison of TnC genes sequences across vertebrates demonstrating a third fish-specific paralog.

The evolutionary history was inferred using the maximum likelihood method based on the JTT-matrix model (27). The bootstrap consensus tree was inferred from 500 replicates with the percentage of replicate trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measure in the number of substitutions per site. The tree is rooted with C. elegans TnC. The analysis involved 37 amino acid sequences of representative vertebrate TnC sequences compiled from NCBI and Ensembl. Evolutionary analyses were conducted in MEGA5 (43).
Figure 5-3  Evolutionary history of TnC1a (cTnC) and TnC1b (ssTnC) amino acid sequences across fish.

The evolutionary history was inferred using A) maximum likelihood method and B) Bayesian analysis, both based on the JTT-matrix model (27). In A) The bootstrap consensus tree was inferred from 500 replicates with the percentage of replicate trees in which the associated taxa clustered together is shown next to the branches. Values represent bootstraps expressed as percentage. In B) the Bayesian analysis was performed over 1x10^6 generations with sampling every 100 generations and a burn-in of 25%. Values represent Bayesian posterior probabilities expressed as percentage. Trees are drawn to scale, with branch lengths indicating the number of substitutions per site and rooted with lamprey cTnC. Evolutionary analyses were conducted in MEGA5 and Topali2.5.
### 5.4.3. Tissue-specific expression patterns

Quantitative real time PCR (qRT-PCR) was used to quantify the tissue-specific gene expression of TnC1a and TnC1b in adult zebrafish at 28°C and adult rainbow trout at 5°C. The dominant transcript measured in zebrafish skeletal mixed muscle with qRT-PCR was TnC1b with minimal amounts of TnC1a (Figure 5-4a). However, in zebrafish cardiac tissue while TnC1a was the dominant paralog (p=0.009), both genes were highly expressed with the expression of TnC1a only 3-fold greater than that of TnC1b (p=0.02) (Figure 5-4b). Meanwhile in rainbow trout cardiac muscle TnC1a was the dominant paralog (Figure 5-4c) while in skeletal muscle TnC1b was almost exclusively expressed (3000-fold greater than cTnC) (Figure 5-4d).

To clarify further the division of expression in cardiac muscle, gene expression was quantified in the two main chambers of the heart, the atrium and ventricle (Figure 5-5). We determined that in adult zebrafish the TnC1a transcript was expressed 150-fold more than TnC1b in the ventricle (p=0.001) when both were normalized to β-actin. Conversely, in the atrium, TnC1b transcript levels were significantly higher (2-fold) than TnC1a (p=0.004).

In 5°C trout, the TnC1a transcript was the dominant paralog in both the atrium and ventricle, though the atrium had only a 23-fold difference in TnC1a relative to TnC1b, while a 360-fold difference in TnC1a relative to TnC1b was observed in the ventricle (Figure 5-5e). In both species β-actin and ef1a transcript levels were comparable in both chambers.

### 5.4.4. Acclimation gene expression patterns

Zebrafish acclimated at colder temperatures (18°C) displayed significantly increased relative expression of TnC1a in both chambers of the heart (Figure 5-5). In the ventricle, TnC1a went from 150-fold higher than TnC1b at 28°C to 300-fold higher in 18°C (p=0.0004). In the atrium, TnC1a went from 2-fold lower than TnC1b at 28°C to 1.5-fold higher (p=0.005) at 18°C. Again, β-actin and ef1a transcript levels did not significantly change with temperature in either chamber.
The profile of both paralogs in either chamber still differed significantly from the 5°C rainbow trout. To clarify whether this difference was simply due to decreased temperature or a phylogenetic difference, rainbow trout were seasonally acclimated to 15°C. While TnC1a was still the dominant isoform in both chambers, TnC1b expression did increase with higher temperature.

![Figure 5-4](image)

**Figure 5-4** Quantitative real time PCR was used to determine tissue specific differences in relative mRNA levels of TnC paralogs in adult zebrafish at 28°C and adult rainbow trout at 5°C.

Values are expressed as mean (n=8) of mRNA levels normalized to β-actin by the ∆∆Ct method (51) and are then expressed as fold-difference from the lower expression isoform. Vertical error bars represent standard error of the mean. Note the scaling of the y-axis in a) skeletal muscle is 4-fold larger than b) cardiac muscle in zebrafish and c) skeletal muscle is 30-fold larger than d) cardiac muscle in trout. * represents p<0.05 (Student’s t-test).
Figure 5-5  Quantitative real time PCR was used to determine tissue specific differences in relative mRNA levels of TnC paralogs in adult zebrafish
A) atrial tissue at 18°C; B) atrial tissue at 28°C; C) ventricular tissue at 18°C; D) ventricular tissue at 28°C; and in adult trout E) atrial tissue at 5°C, F) atrial tissue at 15°C, G) ventricular tissue at 5°C, H) ventricular tissue at 15°C. Values are expressed as mean (n=8) of mRNA levels normalized to the geometric mean of β-actin and ef1a by ∆∆Ct method (51) and then relative as fold difference to ssTnC which was set to 1. Vertical error bars represent the standard error of the mean. * represents p<0.05 (Student’s t-test).
5.5. Discussion

The critical role of troponin C in the regulation of contraction has resulted in this protein being highly conserved across a wide range of species. Previously, the fast skeletal and cardiac TnC genes were thought to be the only paralogs resulting from a genome duplication event prior to the vertebrate radiation (20, 34). However, beyond the typical groupings of cardiac and fast skeletal TnC, zebrafish have two apparent paralogs of what has been annotated as TnC1a (40). In the zebrafish embryo, these two paralogs exhibited a tissue specific pattern of distribution in cardiac and slow skeletal muscle, though functionally either would allow for proper contraction of the heart in embryo (40). These paralogs may have arisen by a tandem gene duplication, in which case they should be on the same chromosome. Alternatively, the paralogs may be the result of the teleost-specific whole genome duplication (32), in which case they should be observed on the homologous chromosomes. Our search of genomic databases revealed clearly that both TnC1a and TnC1b occur in virtually every teleost genome examined to date. Moreover, both paralogs appear on the same chromosome in the limited species where genomic location is available (zebrafish - chromosome 23; stickleback - group XII; medaka - chromosome 7). This suggests that the paralogs arose from a tandem gene duplication that occurred in the common ancestor of the teleosts.

5.5.1. Phylogenetic analysis of TnC

More extensive genome sequence information has become available since previous work on the evolution of troponin C (20). These additional species genomic builds allowed us to identify potential TnC orthologs based on sequence similarity to zebrafish TnC1a and TnC1b from ESTs. Signature residues of the TnC1b clade, which distinguish these sequences from the TnC1a clade (specifically Asp2, Val4, Asn11, Glu14, Glu87/88, Gly108 (Figure 5-1)), appear in all teleosts. We posit that the conservation of all these unique residues in TnC1b representatives indicates that these genes are two functionally divergent paralogs. For all teleosts in which two TnC1 sequences were found, representing species that exist over a broad range of different environmental conditions (cold vs. warm; sedentary vs. active), these residues unique to TnC1b were clearly
identified. While no variation exists in the sequence of the regulatory Ca\textsuperscript{2+}-binding site (site 2) including the six coordinating residues, variation does exist in residue 2, one of four residues implicated in the functional difference between mammalian and fish TnC1a Ca\textsuperscript{2+}-binding affinity (19, 22). With TnC1b resembling mammalian TnC1 (possessing the polar acidic aspartic acid) rather than fish TnC1a (polar uncharged asparagine) in residue 2, a possible difference in Ca\textsuperscript{2+} affinity may exist. Functional characterization of binding affinities of both isoforms in zebrafish will be necessary for the physiological impact of these residue variations to be determined.

Phylogenetic analysis of representative vertebrate species shows the typical division of two clades (TnC2 and TnC1) with the addition of a third fish-specific clade (TnC1b - Figure 5-2). The maximum likelihood phylogeny still reflects the accepted phylogeny of fish species and thus does not reveal anything novel about evolutionary adaptations within fish in TnC gene structure. While TnC1 forms a separate clade from TnC2 supported by higher bootstrap value (98% support for the TnC1-specific clade), the values shown for the separate TnC1 paralogs have relatively low support. Within the clades all bootstrap values demonstrate low support, even between phylogenetically diverse species such as fish and mammals. In this study the TnC1 sequence, as in previous studies, is fundamentally similar across species of teleosts. With fewer species sequences available in previous work, this TnC1 sequence similarity has led to the assumption that the N-terminal sequence is completely conserved (20, 46). Hence low support from bootstrapping is reflective of the high sequence similarity of TnC isoforms and is not likely to be indicative of an inaccurate relationship. As well, Bayesian analysis supported the overall configuration of two independent isoforms in teleosts only (TnC1a and TnC1b) with slightly higher support from posterior probabilities taking into account prior probabilities as well as likelihood of placement. As mentioned above, the conservation of TnC1b-specific residues relative to those in TnC1a across teleosts supports the hypothesis of these two genes being separate paralogs. Several species included in previous fish TnC work (20) had to be excluded from our analyses as the extreme N-terminal (MNDIYKA) in these cTnC sequences, in which several TnC1b/TnC1a distinguishing residues are located (Asp2/Asn2, Val4/Ile4), were lacking.
The evolutionary timing of the appearance of two isoforms of cTnC may be inferred from the phylogenetic relationships. While genetic variability may be generated via point mutations, it is probably large scale genomic duplication events that are primarily responsible for the diversity seen across vertebrate gene families (33). Initially a duplication event provides organisms with two sets of genes presumably carrying redundant functions. Redundant genes are normally lost or reduced to pseudogenes but selective pressures may conserve one of the copies while the other may be relieved from these constraints and rapidly evolve. This could result in each copy assuming only a portion of the functions of the ancestral gene (sub-functionalization) or one copy developing an entirely new function (neo-functionalization) (14). The duplicated TnC genes are thought to have sub-functionalized into TnC2 and TnC1 sometime following the split between urochordates, where a primitive heart appeared, and agnathans with the basic chamber structure of a fish heart (13), where distinct TnC2 and TnC1 appear (20). The TnC1 from lamprey, a representative agnathian, is clearly divergent from all other fish species. A single copy of cTnC appearing in both lamprey and across tetrapods appears to be consistent with the ancestral state.

Basal ray finned fish such as the bichir can help with the understanding if these two genes are products of the teleost-specific genome event. Polypteriformes diverged from teleosts prior to the 3R whole gene duplication event specific to teleosts (7), which occurred approximately 440 MYA (3, 32). While bichir TnC appears to be grouped in with the TnC1b based on the sole sequence available, there is still only one copy of a TnC1-like gene. This bichir TnC1 sequence has some residues matching teleost TnC1a (e.g. Asp2), some matching teleost TnC1b (e.g. Glu87/88) and others matching teleost TnC1b or tetrapod cTnC1 (e.g. Val9). This mix of teleost and tetrapod features seen in the bichir sequence is not exclusive to TnC, but has been seen in other proteins involved with E-C coupling (e.g. RyR (10)). Bichir thus provides an interesting case of the pre-teleost duplication ancestral state, with the current two TnC1 paralogs possibly representing a mixture of the ancestral characteristics. This suggests multiple TnC1 paralogs are teleost-specific, but lacking the availability of characterization of the full genome of a species in the Polypteriformes, such as the bichir or the gar (current genome projects are ongoing such as EnsemblPre! for the spotted gar (*Lepisosteus oculatus*) or Vertebrate TimeCapsule for the grey bichir), this may be a premature statement.
While the maximum likelihood analysis shows both bichir and catshark in basal positions to teleost TnC paralogs, their relationship to each other is not completely clear. Bayesian inference shows bichir in particular to be much more strongly grouped with TnC1b than maximum likelihood does (Figure 5-3b). Bayesian posterior probabilities and bootstrap support values often have observed discrepancies (24) with posterior probabilities tending to be more generous than bootstraps; similar to what is seen in the TnC analysis. Bootstrap support merely helps predict whether the same result would be seen if more data were collected not whether the result is correct. Posterior probabilities refer to the probability that the result is correct given the hypothesis or prior information given. More data (further sequences) will decrease the impact of priors and make Bayesian analysis stronger and allow for possible resolution of basal species. However, until these complete genomic and chromosome location data are determined, the available information is suggestive of these paralogs being the result of an ancestral tandem gene duplication persisting only in teleosts.

Even in a protein as highly conserved as TnC, the presence of three genes in teleosts provides evidence of multiple paralogs in the fish lineage. This demonstrates the complications that the additional gene and genome duplications may bring, especially for research relying on cross-species comparisons for traditional techniques of determining the amount of protein translated such as western-blotting. Non species-specific antibodies are unlikely to account for all isoforms and lead to inaccurate quantification of protein expression. Further, generating antibodies to discern the difference between TnC1a and TnC1b would be difficult due to the high degree of sequence similarity and the low feasibility of finding unique epitopes. In fact, with trout seven Tnl isoforms were all detected by one antibody and even semi-quantitative values could not be inferred (1). The presence of multiple isoforms underlies just one of the issues in accurately quantifying proteins in comparative models and hence made it an unattractive option in our current study. These differences in localization are then solely based on mRNA data due to lack of confidence in accurate quantification of protein translation. This genomic information may provide information about the theoretical status of cellular proteins, but proteomic information describes the actual content, which ultimately determines the phenotype. Changes in the transcription of genes are still key to modifying the proteome and the patterns identified here in zebrafish should be considered representative.
5.5.2. **Paralog localization**

We found differential expression in the transcript levels of TnC1a and TnC1b in the adult zebrafish heart *(Figure 5-4)*. The three different isoforms of TnC (TnC2, TnC1a, TnC1b) found in zebrafish have typically been labeled based on their localization in embryo (40), which in turn reflects the classification of different types of muscle. However, in the adult zebrafish gene expression data provided in this study cardiac muscle does not express one isoform, but rather two. The ratio of TnC1b: TnC1a in zebrafish whole heart is suggestive of atrium to ventricle mass ratio (typically between 0.2 - 0.3 (31). Our data explore this further, showing TnC1b is actually predominant in the atrium whereas TnC1a is predominant in the ventricle, but only in zebrafish at warmer temperatures. Even in trout atrium, TnC1b transcript levels are much more pronounced than in ventricle, especially at warmer temperatures. This differential gene expression in these two chambers may be reflective of the unique contraction of the fish atrium and ventricle relative to mammals (16). In mammals, gene expression of sarcomeric proteins is guided both temporally and spatially to contribute to the overall chamber-specific phenotype. Different subsets (paralogs) of sarcomeric protein expression such as myosin heavy chain and myosin light chain confer differing contractile properties to the atrium and the ventricle. Contractility differences between the atrial and ventricular tissue is in part created by chamber-specific isoforms of sarcomeric proteins. These two fish-specific TnC1 isoforms may be a reflection of the increased need of the specialized teleost atrium in contraction.

5.5.3. **Phylogenetic differential paralog localization**

While the chamber-specific profile in zebrafish suggests chamber-specific contractility differences may be guiding expression patterns of TnC paralogs, we did not see a similar pattern in the cold-water fish measured: rainbow trout in which both chambers predominantly express TnC1a. Although both these fish are teleosts, they have distinct phylogenetic histories with different environmental constraints. The salmonid trout are much larger, more active and live at far colder temperatures than the cyprinid zebrafish. Species have specific expression patterns of certain genes to allow them to be innately suited to their environmental conditions which may not be readily predictable by
only looking at a single variable (i.e. species specific thermal compensation in mitochondrial gene expression (5), hypoxia tolerance contributing to ecological niche specialization between species (11). The preferential expression of TnC1a in trout may be indicative of the role of TnC1a relative to TnC1b. Temperature has been suggested to be a potential reason for increased Ca\(^{2+}\) affinity in fish TnC1a relative to mammals. The increased Ca\(^{2+}\) affinity of trout TnC1a allows for function at colder temperatures (4, 18). However, a relative decreased Ca\(^{2+}\) affinity allows for increased rate of contraction via increased rate of Ca\(^{2+}\) offloading that can occur at warmer temperatures. The obvious difference in functionality for TnC paralogs may be speculated to allow for greater variation in Ca\(^{2+}\) affinity. This variation in Ca\(^{2+}\) affinity between isoforms may be seen in differential temperature responses, with one isoform being able to function at colder temperatures while the other is better suited for warmer temperatures. Therefore, regardless the chamber, TnC1a may be needed for adequate contraction of the heart at lower temperatures (e.g. 5ºC) and thus species that live at colder temperatures may express higher innate levels of this isoform. Further studies will be necessary with more phylogenetically relevant comparisons of fish species to determine whether the differential expression profiles are an evolutionary artefact or a direct adaptation to acclimated temperature.

5.5.4. Temperature specific paralog profiles

To explore further the variation in isoform profiles, both species of fish were exposed to an environmental perturbation to determine if the isoforms were differentially sensitive. Because TnC1a is known to be temperature sensitive as well as the fact that the temperature conditions of the two different study species vary by temperature, fish were subjected to a temperature acclimation regime. Zebrafish were cold acclimated and trout were warm acclimated by 10ºC. This 10ºC difference in temperature was adequate to induce a significant shift in paralog expression in the zebrafish and trout (Figure 5-5). The overall trend of TnC1a being predominant in ventricles and at colder temperatures appears in both species. In zebrafish, both the atrium and ventricle express more TnC1a than TnC1b at 18ºC than at 28ºC. In trout, both chambers still express more TnC1a than TnC1b at 15ºC, but TnC1b expression is much greater at 15ºC than at 5ºC. However,
there appears to be phylogenetically different responses. In zebrafish levels of TnC1a decrease in both chambers with increased temperature, in trout, levels of TnC1a remain relatively constant and only TnC1b fluctuate.

Temperature adaptive shifting in gene expression is the basis for physiological acclimation (41). Interestingly, both the paralog profile and the responses to changing temperature do not match with the two species of fish examined here. This suggests a phylogenetic specific pattern in usage of TnC paralogs, or species-specific regulatory sub-functionalization. While TnC1b levels increase with temperature in the atrium of both species, the scope relative to TnC1a transcript levels is not the same. In fact, the transcript levels of TnC1a vary in zebrafish whereas they remain relatively constant in trout. The evolutionary history of environmental temperature may play a role in how each species remodels the cardiac chambers, or regulates the transcriptional levels of contractile proteins.

5.6. Summary

The regulatory sub-functionalization of TnC isoforms demonstrated in this study lends insight into the consequences of variation to the structure-function of cTn and variation in cardiac muscle contraction. Due to this novel fish-specific paralog, the temperature-dependent phenotype of the fish heart may be more complex than previously thought, especially when coupled to the chamber specific differences in expression and responsiveness to temperature seen in zebrafish. Species-specific variation in transcriptional response of TnC1 paralogs was seen with respect to temperature. Variation in isoforms of sarcomeric proteins influencing the chamber-specific contractile properties may be important in achieving physiological versatility and hence the evolution of specializations in the chambers of the fish heart.
5.7. Acknowledgements

This work was supported by grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Heart and Stroke Foundation of Canada to GFT. GFT is the recipient of a Tier I Canada Research Chair.
5.8. References


5.9. Supplementary

**Figure S1.** Quantitative real time PCR was used to determine tissue specific differences in normalized mRNA levels of TnC paralogs in adult zebrafish at 28°C and adult rainbow trout at 5°C. Values are expressed as mean (n=8) of mRNA levels normalized to β-actin by the ∆∆Ct method (51). Vertical error bars represent standard error of the mean.
Chapter 6.
Structural sub-functionalization of Ca$^{2+}$ binding properties of zebrafish cardiac troponin C paralogs

This chapter is modified from the following original research article in preparation:


My contributions to this study included original concept development and construct design, as well as melting temperatures and data interpretation. This chapter also includes evolutionary analyses of functional divergence and measures of structural frustration performed by me. This chapter was written by me based on the original manuscript written by Charles Stevens to incorporate a more evolutionary perspective.
6.1. Abstract

Zebrafish, as a model for teleost fish, have two paralogous troponin C genes that are expressed differentially in response to temperature acclimation. Upon Ca$^{2+}$ binding, troponin C undergoes a conformational change which exposes a hydrophobic patch that interacts with troponin I, and initiates a series of conformation changes that initiates cardiac muscle contraction. The functional characteristics of the teleost specific troponin C paralogs have not yet been characterized. In this study, we have modeled the structural characteristics using Molecular Dynamics (MD) simulation at 18°C and 28°C and calculated the different Ca$^{2+}$ binding properties between the teleost cardiac (TnC1a/cTnC) and slow-skeletal (TnC1b/ssTnC) paralogs through Potential of Mean Force calculations (PMF). These values are compared with thermodynamic binding properties obtained through Isothermal Titration Calorimetry (ITC).

The modelled structures of each of the paralogs are very similar at each temperature, with the exception of helix C, which flanks the Ca$^{2+}$ binding site. This region in zebrafish contains paralog-specific sequence substitutions that influence protein function. By examining the Ca$^{2+}$ binding interaction through MD simulations, we have isolated the energetic contributions of Ca$^{2+}$ binding from those of the TnC conformational change, which cannot be delineated experimentally. The short timescale of MD precludes the inclusion of the conformational change in the calculation. Therefore the changes in the PMF of the Ca$^{2+}$ interaction can be directly attributed to site II, while the ITC analysis includes the Ca$^{2+}$ binding and conformational change of the TnC molecule. ITC analysis has revealed that TnC1b has higher affinity than TnC1a for Ca$^{2+}$ at both temperatures, while each of the paralogs has increased affinity at 28°C when compared to 18°C. Microsecond timescale simulations have calculated that the cTnC1a paralog is able to transition from the closed to open state more frequently than the TnC1b paralog, an unfavorable transition that would decrease the ITC derived Ca$^{2+}$ affinity, but still increase the Ca$^{2+}$ sensitivity of the myofilament. We propose that the preferential expression of TnC1a at lower temperatures increases myofilament Ca$^{2+}$ sensitivity by this mechanism, despite the lower Ca$^{2+}$ affinity that we have measured by ITC.
6.2. Introduction

Ectothermic species tolerate a range of acute and seasonal temperatures through plasticity in protein function and changes in gene expression (1,2). Central to this tolerance is the maintenance of cardiac function across a range of temperatures. Adaptation to acute temperature occurs more quickly than can be accounted for by transcriptional changes, although greater tolerance can be conferred through preferential expression of paralogs for contractile proteins (1). For example, with prolonged cold exposure expression of hypertrophy and protein synthesis associated genes and those that code for the components of the contractile apparatus are increased (1,3,4). Conservation of ligand binding properties is seen when the function of orthologous enzymes at their typical physiological temperature (5). Thus the basis for understanding the conservation of cardiac contractility at cold temperatures in teleosts lies in the variation of biophysical properties of key proteins in the contractile element that are preferentially expressed at different temperatures.

The critical component in translating Ca\(^{2+}\) transients into contraction of the myofilament is the cardiac troponin complex (cTn). cTn contains three proteins: cTnC (TnC1), a Ca\(^{2+}\) sensor protein; cTnI, an inhibitory protein; and cTnT, which tethers the complex to the remainder of the thin filament by interacting with tropomyosin (6). cTnC is a 2-domain protein with 4 EF-hand motifs designated sites I through IV. Site I does not bind Ca\(^{2+}\), site II is responsible for the Ca\(^{2+}\) sensing function of cTnC, while sites III and IV have a structural role and always bind Ca\(^{2+}\) under physiological conditions (7). The cytosolic concentration of Ca\(^{2+}\) increases from approximately 100 nM during diastole to a maximum of ~1.5 \(\mu\)M in systole (8). As the cytosolic Ca\(^{2+}\) concentration increases, the regulatory site II of cTnC becomes saturated, leading to a conformational change that exposes a hydrophobic patch. This hydrophobic patch then interacts with TnI and allows the TnI inhibitory peptide to withdraw from the actin filament (6). The cTnC-Ca\(^{2+}\) binding interaction also interrupts the TnT-Tropomyosin (Tm) interaction, and permits actin/myosin cross-bridge formation and initiates muscle contraction.

The cardiac contractile element of all species examined is less sensitive to Ca\(^{2+}\)
when the temperature is decreased, an effect that is not observed in skeletal muscle (9,10). However this desensitizing effect is rescued in mammalian cardiomyocytes when 4 amino acid substitutions commonly found in fish (Asn2, Ile28, Gln29, and Asp30 (NIQD) are introduced into the mammalian TnC (11). These 4 substitutions presumably act through an allosteric effect on the ability of site II of TnC to bind Ca$^{2+}$, which allows cardiac contraction to be triggered at lower temperatures (12). TnC is a well-characterized protein in terms of structure and function (13-16) with little variation in secondary or tertiary structures of various TnC1 proteins when similar conditions are considered. The change as a function of the NIQD phenotype is difficult to explain in terms of the isolated N-terminal region of the TnC molecule, as there were no obvious differences between the X-ray derived structure of human cTnC and that with the NIQD substitutions introduced (17).

In trout TnC1, changes in temperature modify structure such that the 7°C form closely resembles that of the mammalian TnC at 30 °C, which suggests that there is a common structure at the respective physiological temperature of trout or mammals (16). The structure of trout TnC was reported to be more open at 7°C than at 30°C based on the A-B inter-helical angle, although this was a very slight change due primarily to changes in the helical bend rather than the exposure of the hydrophobic patch of the protein, as the protein was more compact at 7°C. Functional divergence, which creates variability between proteins, is shaped by biophysical constraints such as stability (18). Perturbation of native structure is important in proteins with different functional configurations (19). An example of this is found through the addition of Ca$^{2+}$ to TnC that causes a slight change in the TnC structure. The inter-helical angle between helices A & B of TnC, and the degree of exposure of the hydrophobic-patch indicate only a subtle difference between the Ca$^{2+}$ bound and Ca$^{2+}$ free forms (13). Increased flexibility of this region has been found in TnC and other Ca$^{2+}$-binding proteins such as parvalbumin. This flexibility increase is achieved through increased reliance on entropy changes associated with Ca$^{2+}$ binding in proteins from cold-adapted teleosts (20). This suggests that subtle changes in conformational flexibility between open and closed states may be linked with variation in Ca$^{2+}$ handling.

In addition to the fish-specific differences in the TnC1a sequence, the fish contractile element also adjusts to temperature change, in part, through the differential expression of a recently discovered slow-skeletal TnC1b (ssTnC) variant (1,21). TnC1b
differs from TnC1a by 18 out of 161 amino acid residues. Interestingly, the TnC1b paralog is missing one of the four residues of the fish-specific NIQD, wherein Asn2 is replaced by aspartic acid, commonly found at that position in mammals (22). The presence of this additional TnC paralog in zebrafish is the result of a tandem gene duplication event. The teleost genome has undergone an additional duplication relative to mammals, which resulted in many redundant duplicate genes. While many duplicate genes are lost through degenerative mutations, in some cases the modification of the ancestral gene function results in a division of labour between the duplicates. This critical process, referred to as sub-functionalization (23), results in retention of functional paralogs in the genome (24).

Sub-functionalization may occur on a tissue-specific level as well as in response to environmental stress such as temperature change. When zebrafish are acclimated to 18°C, the cardiac TnC paralog (TnC1a) is expressed preferentially in both chambers of the heart, while after 28°C acclimation the slow-skeletal paralog (TnC1b) is up-regulated, and becomes the dominantly expressed form in the atrium (1). Both of these paralogs appear to have some overlap in function demonstrated by the ability of the TnC1b protein to rescue a TnC1a knockout in zebrafish embryos (21). The differential expression patterns, even with a mixed paralog expression profile in the atrium, indicate that there may be a functional difference with respect to temperature. The classic model of sub-functionalization, the duplication-degeneration-complementation (DDC) model (25), suggests that modifying the regulatory elements can partition ancestral expression patterns without modifying biochemical function (regulatory sub-functionalization). The continuation of the DDC process can facilitate the emergence of novel or modified functions via sequence substitutions through neutral evolution and still maintain global protein stability (structural sub-functionalization) (26). TnC paralogs in fish provide a clear example of regulatory sub-functionalization (1) but it is unclear as to the nature of structural sub-functionalization, or coding-sequence evolution.

In this study we report the thermodynamic properties of the interaction between Ca²⁺ and each of zebrafish TnC1a and TnC1b at 18°C and 28°C. The TnC-Ca²⁺ interaction was found to be endothermic, which is in agreement with a similar result from human cTnC (27). The interaction between either TnC1a or TnC1b with Ca²⁺ is driven by entropy, and increases with temperature. The strength of the interaction of TnC1b with Ca²⁺ is higher
at both temperatures, as measured by ITC. This is despite very limited structural deviation between the simulation-based structural models of each paralog at each temperature. PMF calculations, which do not consider the energetic component of the structural change, yield similar values, for each paralog at each temperature. This discrepancy is resolved through the use of µs timescale simulations, which calculate that the TnC1a protein more readily transitions from the closed to open state, an unfavorable process that leads to the measurement of lower Ca\textsuperscript{2+} affinity by ITC. We posit that the temperature dependent change in myofilament Ca\textsuperscript{2+} sensitivity that is expected as a function of paralog selection is dictated by the favorability of the TnC conformational change, which transduces the Ca\textsuperscript{2+} binding signal to the myofilament. This supports the necessity of understanding the role of both dynamic conformational flexibility as well as static structures in functional divergence of TnC paralogs.

### 6.3. Experimental Procedures

#### 6.3.1. Homology Modelling and Equilibrium Molecular Dynamics Simulations

The initial models for the TnC constructs were generated using the Swiss-model workspace (28). These models used the NMR-derived Ca\textsuperscript{2+}-bound or Ca\textsuperscript{2+}-free N-terminal domain of the human cTnC structure, PDB:1AP4 or PDB: 1SPY respectively as a template. Models were generated for the N-terminal domain of the cardiac and slow skeletal paralogs of TnC.

The resulting models were equilibrated with GROMACS 4.6.5 (29), using the AMBER99-sb-ILDN (30) force field. The simulation system was defined as a periodic cubic box with 1 nm spacing between the edge of the box and the nearest protein atom. The system was solvated using the TIP3P water model (31), and made neutral by replacing randomly selected water molecules with K\textsuperscript{+} ions. An additional pair of K\textsuperscript{+} and Cl\textsuperscript{-} ions was also added to each simulation system. This was followed by steepest descent energy minimization to a tolerance of 10 kJ mol\textsuperscript{-1} nm\textsuperscript{-1} and conjugate gradient energy minimization.
for 10,000 steps. The minimized system was restrained with 1,000 kJ mol\(^{-1}\) nm\(^{-1}\) absolute position restraints on all of the non-solvent atoms and equilibrated for 1 ns.

The restrained simulations were held at 28\(^\circ\)C or 18\(^\circ\)C using V-rescale temperature coupling (32) and isotropic Berendsen pressure coupling (33) with a \(\tau_T\) of 0.1 and \(\tau_P\) of 1.0, respectively. Interactions were calculated using PME electrostatics (34) and the Verlet cut-off scheme (35). Bond lengths were constrained using the LINCS algorithm (36). Production simulations were done in 5 replicated 100 ns NPT simulations with no position restraints; other parameters were carried forward from the position-restrained simulations. Long timescale simulations were identical, with the exception that the runtime was extended to 1 \(\mu\)s. Models of the TnC/TnI\(_{SW}\) complex were generated as described in Chapter 7, and expanded to include simulations of the TnC1a and TnC1b in complex with the TnI1.1 paralog at 18\(^\circ\)C.

The final models were produced by clustering with the Daura algorithm (37) over the backbone and C-beta atoms of each structure across the five trajectory replicates of each paralog-temperature combination; the middle structure of each of the largest cluster from each mutant simulation were used for further analysis. Protein structure superimpositions were performed with VMD (38), and structural quality assessments were carried out using RAMPAGE (39), QMEAN(40), WHATCHECK (41) and MOLPROBITY (42). Quality statistics may be found in supplementary Table S1. Interhelical angles were measured using interhlx (43) over each replicated simulation. Similarly the hydrophobic solvent accessible surface area (h-sasa) of the TnC molecules were calculated with g_sas over each simulation and averaged (44). The number of hydrogen bonds in TnC were calculated with g_hbond and averaged over each simulation.

6.3.2. Free Energy Calculations

The “protein” was placed in a “cubic” box with dimensions: 6 nm x 6 nm x 15 nm and solvated as before, the net charge was neutralized by replacing randomly selected water molecules with K\(^+\) ions. This was followed by steepest descent energy minimization to a tolerance of 10 kJ mol\(^{-1}\) nm\(^{-1}\) and conjugate gradient energy minimization for 1000
steps. The minimized system was restrained with 1,000 kJ mol\(^{-1}\) nm\(^{-1}\) absolute position restraints on all of the non-solvent atoms and equilibrated for 1 ns.

The initial conformations for umbrella sampling were generated as follows: protein alpha helical alpha carbons were restrained in 3 dimensions with a potential with a force constant of 1000 kJ mol\(^{-1}\) nm\(^{-1}\), and the site II Ca\(^{2+}\) were restrained only in the Y & Z dimensions to permit the movement of the ion. A constraint pulling force was applied in the X dimension at 0.001 nm per ps. Windows for Umbrella sampling were extracted from the resulting trajectory at distance intervals of 0.5 Å between 0 and 1 nm, every 1 Å between 1 nm and 2 nm, and every 2 Å between 2 nm and 5 nm.

### 6.3.3. Umbrella Sampling & WHAM

Umbrella simulations were run as above with no restraining potentials aside from the umbrella potential between the Ca\(^{2+}\) ion and the center of mass of the alpha carbons of the TnC molecule. These simulations used the same parameters as in the pull simulations, with the pull rate set to 0, and were run for 30 ns. Analysis was performed using g_wham and the first 5 ns of each sampling window simulation were discarded, errors were estimated with 10,000 bootstraps of the WHAM calculation.

### 6.3.4. Protein Expression & Purification

Synthetic, codon-optimized genes (GeneArt) encoding zebrafish TnC1a (gi 28822163) & TnC1b (gi 50344824) were cloned into the pET-21a(+) expression vector (Novagen), and the codon corresponding to residue 90 was replaced with a stop codon using the Phusion site-directed mutagenesis protocol to produce the N-terminal fragment of each TnC1 sequence (Thermo). The resulting constructs were verified by DNA sequencing and transformed into the BL21(DE3) expression host E. coli strain.

Overnight cultures were diluted 20 fold into LB media and incubated at 37°C shaken at 250 rpm. Cultures were induced at OD\(_{600}\)=0.8 with a final concentration of 1 mM
IPTG followed by 3 h of growth. Cells were harvested by centrifugation, re-suspended in 20 mL of 50 mM Tris-HCl pH 8.0, 5 mM EDTA, 1 mM PMSF and one complete protease inhibitor tablet (Roche). The cells were lysed by sonication on ice at 80% amplitude using ten 30-second pulses at 30-second intervals. Lysate was clarified by centrifugation at 30,000 x g for 30 min and the supernatant was applied to a fast-flow DEAE column (Amersham) equilibrated with 50 mM Tris pH 8.0, 5 mM EDTA, 1 mM DTT and eluted with a 180 mL gradient between the equilibration buffer and the high-salt buffer comprised of equilibration and 550 mM NaCl. Fractions containing the TnC protein were pooled and concentrated to 5 mL using an Amicon centrifugal concentrator with a MWCO of 3,000 Da (Millipore). The concentrated protein was applied to a HiPrep 26/60 Sephacryl S-100 column equilibrated with 50 mM Tris pH 8.0, 100 mM NaCl and 1 mM DTT. Fractions containing the TnC protein were pooled, concentrated and stored at -80°C.

6.3.5. Melting Point Determination

Protein solutions in the apo state were dialyzed 4 times against 2 L of MT buffer (150 mM KCl, 10 mM HEPES pH 7.5, 3 mM MgCl2, 2 mM EGTA). Calcium-saturated TnC was dialyzed and diluted similarly with MT buffer supplemented with 3 mM of CaCl2.

Each melting temperature replicate contained 3 mg/ml of TnC and 2.5 µL of 100X diluted sypro orange. Temperatures were increased from 4°C to 95°C at 5 second intervals on a BioRad CFX96 Touch Real-Time PCR System (BioRad). The melting temperature of the protein was determined at the peak of the first derivative curve representing the midpoint of the unfolding transition.

6.3.6. Isothermal Titration Calorimetry

Protein solutions were dialyzed 3 times against 2 L of ITC Buffer (50 mM HEPES pH 7.2, 150 mM KCl) and diluted to a final concentration of 200 µM. The first two buffers contained 15 mM, β-mercaptoethanol (β-ME) and the last contained 2 mM β-ME. Initial buffer contained 2 mM ethylenediaminetetraacetic acid (EDTA) to generate TnC in the
apo-state with two subsequent exchanges to remove the EDTA. This was verified by the Edelhoch method (45) using an extinction coefficient of 1490 and a MW 10.1 kDa (looked up on protparam). Ca\textsuperscript{2+} solution was prepared in the dialysis buffer from the final exchange to a final Ca\textsuperscript{2+} concentration of 4 mM. Ca\textsuperscript{2+} was titrated into the protein solution in a series of 19 injections of 2 µL (the first being a dummy injection of 0.4 µL), 2 min apart with a stirring speed of 1000 rpm. These experiments were repeated at 18°C and 28°C, and the last 3 data points were subtracted from the Ca\textsuperscript{2+} - protein titrations to account for interaction of Ca\textsuperscript{2+} with the buffer and non-specific interactions with the protein. Origin 8.0 was used to analyze the data.

6.3.7. Functional divergence of amino acids

Functional divergence comparing the predicted proteins of genes in each clade of interest was calculated using the statistical framework provided in the GU99 method of DIVERGE 3.0 (46). Type I and type II divergence were compared between clusters of TnC (1a vs 1b)). Coefficients of $\theta_I$ and $\theta_{II}$ significantly $>$0 demonstrate divergence between clusters. For type I functional divergence, to reject the null hypothesis ($\delta \theta =$0), we compared two times the standard error of theta and a LRT (47). For type II functional divergence, pairs with $\theta_{II}$ values greater than 0 after subtracting two times the standard error were considered significant.

6.3.8. Analysis of local frustration

The single residue level frustration (SRLF) was computed from the ‘frustratometer’ server (http://www.frustratometer.tk) utilizing the mutational frustration index on the minimized models generated for TnC1a and TnC1b at 18°C and 28°C. This index is estimated from the energy of a protein complex compared to the energies of a set of ‘decoy’ states. The local SRLF index for a particular amino acid was defined as a Z-score of native state energy compared to the N decoys. Residues with a Z-score $>$1 are defined as stabilizing or minimally frustrated, while residues with a Z-score of $<$-1 are defined as
destabilizing or highly frustrated (48). Between these limits a residue is described as neutral.

6.4. Results

6.4.1. Homology Models

After 100 ns of simulation, each of the modeled systems had diverged from the starting coordinates substantially (Figure S1). The middle structure of the largest cluster for each paralog at each temperature was selected as representative for further analysis. The quality indicators for the representative structures used in the PMF calculations may be found in (Table S1).

The structures of the equilibrated homology models are very similar to each other (Figure 6-1), with the exception of the relative orientation and stability of helix C. Helix C is located immediately following Ca²⁺ binding site II and changes in this helix may affect the Ca²⁺ interaction. Helix C contains two sequence substitutions between the TnC paralogs, proline 54 in TnC1a is replaced with a glutamine in TnC1b, and isoleucine 60 in TnC1a is replaced with a valine in TnC1b. The secondary structure, as calculated by STRIDE, is variable between paralog temperature combinations, with helical lengths that vary by up to 4 residues in the case of Helix A (Figure 6-2). Despite the fact that there was no change in the length of helix C for each paralog with respect to temperature, the orientation of that helix was the greatest difference between the 18°C and 28°C structures. In TnC1a at 18°C helix C is nearly anti-parallel to helix D at 153°, while in the other representative structures helix C and D are at angles of 128°, 123° & 112° for TnC1b at 28°C, TnC1b at 18°C and TnC1a at 28°C, respectively.
Figure 6-1  Superimposed structures of the equilibrated representative structures after Daura clustering using backbone and beta carbon atoms over the duration of 100 ns equilibration trajectories. Helices are labeled A, B, C and D. To the left are the cardiac paralog at 28°C (dark blue) and 18°C (light blue) and to the right are the slow-skeletal paralog structures at 28°C (dark red) and 18°C (salmon). The differences between these structures are minimal, but most pronounced in the orientation of helix C relative to the remainder of the protein.

Figure 6-2  Sequence alignment of ZF TnC1a and TnC1b, highlighted to visualize the changes in secondary structure after equilibration at each temperature and energy minimization. The most salient change is in helix C, and is likely due to the temperature dependent effect of the P55Q, I61V and E75D substitutions on the stability of helix C and its interaction with Helix D.
Inter-helical angles were monitored over the course of each simulation; none of the structures reached the open state, defined by an AB inter-helical angle below 90° (Figure 6-3). The average AB-inter-helical angles are reported in Table 1. During the 5 replicated 1 µs simulations the TnC structures at 28°C each had several frames in an intermediate state, 7 frames were below 105° in the Ca²⁺-free, whereas 359 frames were below 105° in the Ca²⁺-bound simulations. There was only a single frame that contained a structure of the TnC1b molecule with an AB interhelical angle below 105° (Table S2).

Table 6-1  Average AB inter-helical angle over 5 replicated 1 µs simulations of Ca²⁺ bound and Ca²⁺ free TnC1a (cTnC) and TnC1b (ssTnC).

<table>
<thead>
<tr>
<th></th>
<th>18°C (Ca²⁺ bound)</th>
<th>28°C (Ca²⁺ bound)</th>
<th>18°C (Ca²⁺ free)</th>
<th>28°C (Ca²⁺ free)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TnC1a</td>
<td>140° ± 7</td>
<td>125° ± 9</td>
<td>134° ± 5</td>
<td>131° ± 9</td>
</tr>
<tr>
<td>TnC1b</td>
<td>132° ± 9</td>
<td>133° ± 6</td>
<td>134° ± 5</td>
<td>132° ± 8</td>
</tr>
</tbody>
</table>

Figure 6-3  Superimposition of snapshots of TnC1 demonstrated inter-helical angles
(A) Superimposition of snapshots from the 1µs simulation of TnC1a at 28°C, helices are labeled N, A, B, C and D. The most open conformation from this simulation is shown in blue, while the most closed conformation is shown in white. (B) Angles are drawn onto the superimposed and isolated A and B helices. (C) The Helices are superimposed with equivalent structures from the NMR-derived human cTnC structures in the Ca²⁺-free, closed form (PDB:1SPY) shown in red, and the open conformation in complex with the TnI switch peptide (PDB:1MXL) shown in green.
When the h-sasa of each structure at each temperature was calculated, there were only slight differences as a function of temperature or sequence substitution; however when the h-sasas are calculated for either the TnC1a or TnC1b molecule isolated from the TnC/TnI complex and therefore in the open conformation, the average h-sasa of TnC1a at 28ºC increased by 4.9 nm² while the average h-sasa of TnC1b at 28ºC increased by 4.1 nm² as a result of the closed to open transition. At 18ºC the closed to open transition increased the h-sasa by 4.2 nm² and 4.6 nm² for TnC1a and TnC1b, respectively. This is in contrast to the effect of Ca²⁺ binding, which decreased the h-sasa of TnC1a at 18º and 28ºC by 0.2 nm² and 0.6 nm², respectively and decreased the h-sasa of TnC1b at 28ºC by 0.7 nm² but increased the h-sasa of TnC1b at 18ºC by 0.2 nm² (Table 6-2).

Table 6-2  
Average hydrophobic solvent accessible surface area (nm) over the final 50 ns of 5 replicated 100 ns simulations.

<table>
<thead>
<tr>
<th></th>
<th>18ºC (Ca²⁺ bound)</th>
<th>28ºC (Ca²⁺ bound)</th>
<th>18ºC (Ca²⁺ free)</th>
<th>28ºC (Ca²⁺ free)</th>
<th>18ºC (Open)</th>
<th>28ºC (Open)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TnC1a</td>
<td>22.3 ± 0.4</td>
<td>21.8 ± 0.5</td>
<td>22.6 ± 0.6</td>
<td>22.4 ± 0.7</td>
<td>26.5 ± 0.9</td>
<td>26.8 ± 0.5</td>
</tr>
<tr>
<td>TnC1b</td>
<td>21.0 ± 0.5</td>
<td>20.8 ± 0.7</td>
<td>20.8 ± 0.8</td>
<td>21.5 ± 0.5</td>
<td>25.6 ± 0.9</td>
<td>25.0 ± 0.8</td>
</tr>
</tbody>
</table>

The flexibility of each of the paralog-temperature combinations is similar (Figure 6-4), represented by the root mean squared fluctuation of each Cα atom, sampled over the final 10 ns of each of the simulations and averaged over 5 replicated simulations, as are the number of hydrogen bonds, which are identical for the Ca²⁺-free TnC molecules, but increased by approximately 1 hydrogen bond in each paralog-temperature combination in the Ca²⁺ bound state (Table 6-3).

Table 6-3  
Average number of hydrogen bonds in each structure over the final 100 ns of 5 replicated 1 µs simulations for Ca²⁺ bound and Ca²⁺ free models of TnC1a and TnC1b

<table>
<thead>
<tr>
<th></th>
<th>18ºC (Ca²⁺ bound)</th>
<th>28ºC (Ca²⁺ bound)</th>
<th>18ºC (Ca²⁺ free)</th>
<th>28ºC (Ca²⁺ free)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TnC1a</td>
<td>62.0 ± 1.6</td>
<td>61.0 ± 1.3</td>
<td>60.9 ± 0.6</td>
<td>60.6 ± 1.0</td>
</tr>
<tr>
<td>TnC1b</td>
<td>61.8 ± 0.8</td>
<td>61.3 ± 1.6</td>
<td>60.9 ± 1.4</td>
<td>60.6 ± 1.2</td>
</tr>
</tbody>
</table>
Figure 6-4  Average root mean squared fluctuation per residue. Mean and SEM (n=5) are plotted for the final 50 ns of each of 100 ns simulations for the paralog-temperature combinations. The protein is most flexible at the termini and EF-hand loops. There is limited variability in RMSF between paralog-temperature combinations, which suggests that the sequence differences do not greatly influence the protein flexibility. There is a slight effect of temperature on flexibility in the loop regions between helices A and B, as well as the loop between helices C and D.

Figure 6-5  Calcium coordination distances for each paralog-temperature combination. Residues that may offer substitute ligands are also included. The distance values are similar for each paralog at each temperature. Ser69-OG is slightly closer to the Ca$^{2+}$ ion in the TnC1a (cTnC) simulations, while asp 73 is slightly closer to the Ca$^{2+}$ ion in the TnC1b (ssTnC) simulations.
The coordination distances for the Ca\textsuperscript{2+} ions were measured for each of the side chain ligand atoms, as well as any nearby potential substitutes. These values are plotted in Figure 6-5. There was very little difference between paralog-temperature combinations, particularly in the distances between ligand atoms that are close enough to be involved in coordination of the Ca\textsuperscript{2+} ion.

6.4.2. Free Energy Calculations

The change in free energy for each of the paralog temperature combinations was determined using PMF calculations (Figure 6-6); there was good coverage over the range of the center of mass distance that was used to calculate the free energy of Ca\textsuperscript{2+} interaction for the TnC paralogs at each temperature. The umbrella potential is plotted in Figure 6-7, the free energy differences between these plots show that at 28°C there was little difference between the paralogs, while at 18°C, TnC1a has a higher $\Delta G$ of Ca\textsuperscript{2+} interaction than TnC1b. The calculated free energy values are listed in Table 6-4.

**Figure 6-6**  PMF reaction Coordinate.
The Ca\textsuperscript{2+} ion is removed from the site II binding loop along the X axis. Vertical lines indicate the frequency of PMF sampling windows over the reaction coordinate. Sampling windws were arranged at 0.5 Å below 1.0 nm distance, 1 Å between 1.0 and 2.0 nm, and 2 Å between 2.0 and 5.0 nm.
Figure 6-7  WHAM derived Umbrella Potentials for each Temperature-paralog combination.

The plots for each paralog appear to overlap at each temperature; however, the values at low COM distance reach minima at different points. This indicates that the ssTnC-Ca\(^{2+}\) interaction at 18°C (shown in green) is less favorable than the others, which are approximately equal. At 28°C the $\Delta G$ of Ca\(^{2+}\) interaction for cTnC and ssTnC were $-55.5 \pm 4.1$ kJ mol\(^{-1}\) and $-58.0 \pm 2.7$ kJ mol\(^{-1}\), respectively. At 18°C, the $\Delta G$ of Ca\(^{2+}\) interaction for cTnC was $-51.4 \pm 3.6$ kJ mol\(^{-1}\) and ssTnC was at $-45.5 \pm 3.0$ kJ mol\(^{-1}\).

Table 6-4  Free energy of Ca\(^{2+}\) interaction calculated through Potential of Mean Force (kJ/mol) Errors are the standard deviation from 10,000 bootstraps

<table>
<thead>
<tr>
<th></th>
<th>18°C (kJ/mol)</th>
<th>28°C (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TnC1a</td>
<td>-51.4 ± 3.6</td>
<td>-55.5 ± 4.1</td>
</tr>
<tr>
<td>TnC1b</td>
<td>-45.5 ± 3.0</td>
<td>-58.0 ± 2.7</td>
</tr>
</tbody>
</table>
6.4.3. Melting Point determination

The Tm of the Apo-TnC1a and Apo-TnC1b were 70.1°C and 65.9°C, respectively. The Tm of each of the Ca$^{2+}$ bound forms of the proteins were approximately 15°C higher in both cases at 84.9°C and 82.3°C for TnC1a and TnC1b, respectively. Due to the exposure of the hydrophobic TnI-binding patch on the surface of TnC as a result of Ca$^{2+}$ interaction, the fluorescence measurements were substantially higher at all temperatures in the Ca$^{2+}$-saturated experiments, which may have right-shifted the inflection point of the melt-curve and inflated the Tm values.

6.4.4. Isothermal Titration Calorimetry

The interaction between each of the N-TnC constructs and Ca$^{2+}$ were endothermic in nature. In each case the stoichiometric ratio of Ca$^{2+}$ binding was approximately 1, indicating that the regulatory site II EF-hand was exclusively titrated during these experiments. The $\Delta S$ values were higher for TnC1a at 28°C than 18°C, which were each greater than $\Delta S$ values for TnC1b, which did not differ significantly with temperature. The $\Delta H$ values were greatest for the TnC1a paralog, Ca$^{2+}$ binding to TnC1a at 28°C yielded a greater $\Delta H$ in response to increased temperature, while the $\Delta H$ values reported for TnC1b did not differ significantly with temperature. The $\Delta G$ values were most favourable for TnC1b at 28°C, followed by TnC1b at 18°C and TnC1a at 28°C, which did not differ significantly; the $\Delta G$ value for TnC1a at 18°C was the least favourable. The thermodynamic parameters are listed in Table 6-5.
Table 6-5  Thermodynamic Parameters derived from ITC for each paralog-temperature combination (n=3). For each parameter mutant-temperature conditions not having the same superscripted letter are significantly different p<0.05.

<table>
<thead>
<tr>
<th></th>
<th>cTnC (18°C)</th>
<th>cTnC (28°C)</th>
<th>ssTnC (18°C)</th>
<th>ssTnC (28°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kd (µM)</td>
<td>19.0 ± 1.2A</td>
<td>16.7 ± 0.8II</td>
<td>11.5 ± 0.7IC</td>
<td>10.2 ± 0.8ID</td>
</tr>
<tr>
<td>ΔS (J/mol)</td>
<td>153.9 ± 2.1B</td>
<td>158.6 ± 5.0A</td>
<td>147.7 ± 2.9C</td>
<td>148.1 ± 2.1IC</td>
</tr>
<tr>
<td>ΔH (kJ/mol)</td>
<td>18.5 ± 0.5B</td>
<td>20.2 ± 1.4A</td>
<td>15.4 ± 0.8IC</td>
<td>15.9 ± 0.5IC</td>
</tr>
<tr>
<td>ΔG (kJ/mol)</td>
<td>-26.4 ± 0.2A</td>
<td>-27.6 ± 0.1B</td>
<td>-27.5 ± 0.1B</td>
<td>-28.8 ± 0.2C</td>
</tr>
</tbody>
</table>

The TnC1a paralog was more sensitive to temperature. The Kd value decreased from 19.0 µM to 16.7 µM with the 10°C temperature increase, while the Kd value for TnC1b decreased from 11.5 µM to 10.2 µM as temperature increased from 18°C to 28°C.

6.4.5. Functionally divergent sites

Residues that are highly conserved in one paralog but variable in another are represented by Type I functional divergence. No type I divergence was detected in N-TnC. Amino acids that are highly conserved within each paralog sub-clade but vary between paralog groups are represented by Type II functional divergence. All sites measured as functionally divergent across teleost phylogeny are located in N-helix, or DE linker in N-TnC (residue 2, 4, 11, 87, 88) (Figure 6-8). There are several other sites that vary between ZF TnC1a and TnC1b, but these are species-specific to ZF and not detected across phylogeny.
Figure 6-8  Type II functionally divergent sites in ZF TnC1 paralogs as detected by DIVERGE.

Superimposed modelled structures of ZF TnC1a (blue) and TnC1b (red) at 18ºC demonstrate functionally divergent residues are located in N-helix and DE-linker regions of the protein.

6.4.6. Analysis of local frustration

The principle of minimum frustration is key to the maintenance of native structure but residual energetic frustration is important to protein function (48). Changes in energetic frustration are linked to functional regions of proteins, or sites that can be linked to functional changes. The SRLF was computed for functionally divergent residues as well as species-specific residues contributing to variation in helix C (Table 6-6, Figure 6-9). Of the functionally divergent residues, residue 2 is more highly frustrated in TnC1a than TnC1b at both temperatures. Residue 4 and residue 88 both show higher SRLF scores in TnC1a at both temperatures, which estimates lower frustration relative to TnC1b. Residue 54 has a neutral Z-score in both TnC1a and TnC1b at 18ºC, but at 28ºC P54 in TnC1a is highly frustrated while Q54 is still neutral in TnC1b. Residue 61 was estimated to be minimally frustrated in both paralogs, but higher Z-scores at both temperatures suggest more a stable conformation in TnC1a. Residue 75 is highly frustrated in all conditions,
which suggests of a greater role in conformational change that is consistent in both paralogs regardless of the amino acid at this site.

Table 6-6  Single Residue Level Frustration Index

<table>
<thead>
<tr>
<th>Residue</th>
<th>2</th>
<th>4</th>
<th>11</th>
<th>54</th>
<th>61</th>
<th>75</th>
<th>87</th>
<th>88</th>
</tr>
</thead>
<tbody>
<tr>
<td>TnC1a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18°C</td>
<td>-1.203</td>
<td>1.876</td>
<td>-0.842</td>
<td>0.383</td>
<td>1.842</td>
<td>-1.052</td>
<td>-0.017</td>
<td>1.454</td>
</tr>
<tr>
<td>28°C</td>
<td>-1.156</td>
<td>2.333</td>
<td>-0.780</td>
<td>-1.450</td>
<td>1.651</td>
<td>-1.124</td>
<td>-0.151</td>
<td>1.303</td>
</tr>
<tr>
<td>TnC1b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18°C</td>
<td>-0.908</td>
<td>1.386</td>
<td>-0.765</td>
<td>0.073</td>
<td>1.419</td>
<td>-1.096</td>
<td>0.426</td>
<td>0.706</td>
</tr>
<tr>
<td>28°C</td>
<td>-0.863</td>
<td>1.376</td>
<td>-0.742</td>
<td>0.117</td>
<td>1.429</td>
<td>-1.104</td>
<td>0.390</td>
<td>0.723</td>
</tr>
</tbody>
</table>

Figure 6-9  Variation in Helix C and D of ZF TnC1 paralogs.

Superimposed modelled structures of ZF TnC1a (blue) and TnC1b (red) at 18°C demonstrate three sequence substitutions result in changes in length and orientation of helix C.
6.5. Discussion

Functional divergence incorporates both regulatory and structural sub-functionalization. Examining the effect of sequence substitutions on Ca\(^{2+}\) interaction in the presence and absence of the conformational change has shed mechanistic insight into the role of these variations in the function of TnC. The increased Ca\(^{2+}\) sensitivity of cardiomyocytes is seen in teleosts relative to mammals is attributed to fish specific NIQD molecular phenotype (49,50). The explanation for the increase in sensitivity was that these sequence substitutions either allosterically influence the affinity of site II for Ca\(^{2+}\), or lower the energy barrier associated with the conformational change that follows Ca\(^{2+}\) binding, thereby enhancing the contraction of the myofilament. The energetics of the conformational change can be considered as a trade-off between the cost of the Ca\(^{2+}\)-binding induced strain on the structure of TnC and the cost of exposing the hydrophobic patch of the TnC molecule to aqueous solvent (51). For retention of functional fish specific TnC paralogs, structural stability must be balanced with sufficient flexibility for the conformational changes associated with Ca\(^{2+}\) binding (52). In this study we combine a computational framework to predict the energy changes with Ca\(^{2+}\) binding with an experimental measure of thermal energy release with Ca\(^{2+}\) binding in the same proteins. TnC1b has higher affinity than TnC1a for Ca\(^{2+}\) at both temperatures as measured by ITC, while MD reveals subtle calculated structural changes in each of the paralogs. TnC paralogs in zebrafish thus provide a model of the interplay of structural stability and conformational flexibility in sub-functionalization.

Biochemical functions are often conserved and tertiary structures are usually highly similar even when sequences have diverged (53). Proteins are generally robust to single-site mutations and need to demonstrate sufficient plasticity in native folding patterns to maintain overall structural stability (54). Evolutionary conservation of a site is correlated with the relative solvent accessibility of a site in tertiary structure and presence of hydrogen bonds, both of which contribute to the overall stability of the protein (55-57). Evolutionary constraints select for proteins stable enough to perform function reliably, but this must be balanced with the need for conformational flexibility to allow for binding events (58). Increased stability of TnC, conferred by a stronger hydrophobic core, or antiparallel
stacking of helices to mutually stabilize helical dipoles will preferentially stabilize the closed conformation and decrease the affinity for Ca\(^{2+}\) by increasing the energetic cost of the conformational change. A more stable closed conformation relative to the open conformation will increase the energetic cost of the closed-open conformational change and decrease the affinity for Ca\(^{2+}\). Changes in secondary structure have been noted in response to Ca\(^{2+}\) binding of other TnC proteins through the use of CD spectroscopy, such that greater helical content in the Ca\(^{2+}\) bound form conferred greater stability and increased the favourability of the Ca\(^{2+}\) interaction (59,60). The melting temperatures of Apo and Ca\(^{2+}\) bound forms vary by 15ºC as a function of Ca\(^{2+}\) interaction in both TnC1a and TnC1b, indicative of similar relative stability. Paradoxically TnC1a, which is dominantly expressed on the transcriptional level at lower temperatures, has a higher melting point than TnC1b despite RMSF values that show that flexibility of the paralogs are approximately equal (Figure 6-4). The MD simulations predict that there is one additional hydrogen bond on average in the Ca\(^{2+}\) bound state, when compared to the Ca\(^{2+}\) free state, however the total number of h-bonds were similar between paralogs at each temperature (Table 6-3). The h-sasa upon Ca\(^{2+}\) binding decreased for each paralog-temperature combination, with the exception of ssTnC at 18ºC, which suggests that any stabilizing effect of Ca\(^{2+}\) binding was less pronounced in that condition.

TnC is a highly conserved protein across phylogeny, with few sequence changes maintained in any vertebrate ortholog. Overall, there is very little difference in the equilibrated homology models of these two paralogs. There are only 18 substitutions out of 161 aa between the TnC1a and TnC1b in zebrafish and across species TnC1a and TnC1b share 87-89% identity. None of the variation between TnC1a and TnC1b occurs in Ca\(^{2+}\) coordinating residues of site II (Figure 6-5). Increased Ca\(^{2+}\) binding affinity is a typical functional change in TnC, which in turn determines the Ca\(^{2+}\) sensitivity of the entire sarcomere. In TnC1, Ca\(^{2+}\) affinity refers directly to the interaction between site II and Ca\(^{2+}\). The association of Ca\(^{2+}\) to this motif is not only dependent on the coordinating residues but also the overall conformation of the protein. Substitutions in the EF-hand motif that directly influence Ca\(^{2+}\) affinity are not the only way to change the function of Tn. Ca\(^{2+}\) sensitivity refers to the propensity of a muscle contraction at a particular concentration of Ca\(^{2+}\). Drugs and mutations that stabilize the open conformation of N-TnC produce an apparent increase in affinity that does not alter the coordination of Ca\(^{2+}\) by site II, but rather
the increases the favourability of the conformational change, which is driven by the release of the conformational strain on the tertiary structure (52). Conversely, sequence substitutions that directly increase the relative stability of the closed conformation over the open form produces an apparent decrease in the \( \text{Ca}^{2+} \) affinity of site II of TnC. This apparent change in affinity occurs despite the unchanged coordination of the \( \text{Ca}^{2+} \) ion by site II. Residue variation in TnC1a may increase molecular flexibility facilitating the protein to be activated by lower concentrations of \( \text{Ca}^{2+} \) (61). Even with minimal structural change, functional variation can occur in response to sequence substitutions in TnC. Changes in the stability of the protein may be relatively small in comparison to variations in the conformational flexibility of the protein in the transition between open and closed conformations.

Calcium affinity of TnC has been difficult to reliably measure due to the thermodynamic contributions of the conformational change. MD-derived \( \Delta G \) values represent the change in affinity of site II as a function of structural changes induced by sequence substitution and equilibration at each temperature, but do not account for the thermodynamic consequences of the open/closed conformational change. The MD simulations presented here indicate that the \( \Delta G \) of \( \text{Ca}^{2+} \) binding is similar for both paralogs at 28ºC, and TnC1a at 18ºC while the \( \Delta G \) of \( \text{Ca}^{2+} \) binding for TnC1b at 18ºC is less favourable. The timescale of these simulations are inadequate to sample the conformational change of the protein, and therefore cannot include its energetic contribution to \( \text{Ca}^{2+} \) binding. The process of \( \text{Ca}^{2+} \) binding to \( \text{Ca}^{2+} \) sensing EF-hand proteins has been described as a balance between conformational strain and conformational change (51). The regulatory N-TnC fluctuates between open and closed conformation and facilitates the likelihood of \( \text{Ca}^{2+} \) binding. \( \text{Ca}^{2+} \) binding precipitates a shift in the equilibrium between the primed and open states, with the proportion shifting from 0% open in the apo state to between 20% and 27% in the \( \text{Ca}^{2+} \) bound state (62,63). The free energy cost of this transition has been estimated at 8 kcal/mol (33.5 kJ/mol) for \( \text{Ca}^{2+} \) bound human wild-type TnC through the use of long timescale simulations (64). This can confound ITC measurements, as the energetics of the shift in equilibrium between open and closed states cannot be experimentally decoupled from the \( \text{Ca}^{2+} \) binding interaction. The opening of the N-cTnC molecule exposes a hydrophobic patch, an energetically unfavourable
process that is well documented for EF-hand proteins, which is offset by the relief of conformational strain that builds as a function of Ca\textsuperscript{2+} binding (51).

The structural reason for the observed change in the ITC $\Delta G$ of Ca\textsuperscript{2+} binding to TnC1b at 28\textdegree C may be due to the predicted change in the length and orientation of helix C, which is adjacent to site II, which alters the C/D helical interface (Figure 6-2). This helix is much shorter in the representative structure of TnC1b at 28\textdegree C than the others, suggesting that the stability of this helix may be compromised at 28\textdegree C while under the conformational strain induced by Ca\textsuperscript{2+} binding. Helices C and D contain three sequence substitutions between TnC1a and TnC1b. Residues 54, 61, and 75 are Proline, Isoleucine and Glutamic acid respectively in TnC1a are replaced by Glutamine, Valine and Aspartic acid in TnC1b. The I61Q substitution in human cTnC has been shown to affect the stability of the helical packing (65). The P54Q substitution in particular may increase affinity by destabilizing the closed conformation in TnC1b, and relieve the conformational strain through the shortening of Helix C. Changes in energetic frustration of individual residues can facilitate movement (58). While the latter two residues substitution display minor changes in frustration, variation in SLRF index at residue 54 between paralogs is seen at 28\textdegree C (Table 6-6). The change in SLRF of TnC1a at 28\textdegree C to more highly frustrated state is favourable for conformational flexibility or increased likelihood to transition to the open state without the shortening of Helix C. This subtle variation between paralogs in helix C is enough to cause differential response to conformational strain resulting in changes to the energetics profile.

The mammalian cardiac contractile element is less sensitive to Ca\textsuperscript{2+} when the temperature is decreased (10) therefore it is understandable that a higher relative TnC affinity is needed in fish to allow for proper cardiac function in colder environments (49,66). Thermal stability of a protein is related to the need to maintain the appropriate conformational flexibility at physiological temperatures, allowing the conformational changes that accompany binding and catalysis to occur at appropriate rates (56). Increased temperature is commonly used in MD simulations to accelerate the rate at which conformational space can be sampled. Increased temperature of the system enables the traversal of energetic barriers that are below the temperature of the system. This principle, when applied to the TnC molecule suggests that in order to maintain functional
equivalence the energy barrier that separates the open and closed conformations of the TnC protein should be lower for TnC1a than TnC1b as TnC1a is dominantly expressed at lower temperatures (1) Similarly, the energy barrier for the TnC1b conformational transition should be higher, to decrease Ca$^{2+}$ sensitivity at higher temperatures. This is correlated with the size of the hydrophobic patch that is exposed for TnC1b and TnC1a. At 28ºC, the h-sasa is 1.8 nm$^2$ larger on average for TnC1a. However the transition between the closed and open states saw the h-sasa for TnC1b increase by 4.6 m$^2$ (at 18ºC) or 4.1 nm$^2$ (at 28ºC), but only by 4.2 m$^2$ (at 18ºC) or 4.9 nm$^2$ (at 28ºC) for TnC1a, which suggests that there may be a lower barrier to the transition for TnC1a.

Given that higher relative Ca$^{2+}$ sensitivity is expected at lower temperatures (67), and TnC1a is dominantly expressed at low temperatures (1), we expect that TnC1a should have higher affinity for Ca$^{2+}$. This is not what is found when the affinities are measured by ITC where TnC1b has higher affinity in all cases, or by PMF analysis, where TnC1b has lower Ca$^{2+}$ affinity at 18ºC but TnC1b Ca$^{2+}$ affinity is the same as TnC1a at 28ºC. The timescale of the PMF simulations are inadequate to sample the conformational change of the protein, and therefore cannot include its energetic contribution. Monitoring the A/B interhelical angle over 1µs simulations has shown that the Ca$^{2+}$ bound TnC1a molecule is more amenable to conformational change than TnC1b. The interhelical angle did not fall below 90º, the criterion for the open state that has been used in similar simulations (64), but it did not remain completely closed for as much of the simulation as TnC1b, and was more frequently observed with lower A/B interhelical angle values (Table S2). From this we infer that TnC1a has a less stable closed conformation than TnC1b, and is therefore less able to tolerate the conformational strain induced by Ca$^{2+}$ binding. The TnC1a is more likely to transition between the closed and open states, which should create higher myofilament Ca$^{2+}$ sensitivity, while still maintaining lower Ca$^{2+}$ affinity we have observed by ITC. The TnC1b molecule has higher affinity for Ca$^{2+}$, as measured by ITC, which is conferred by the more stable closed conformation and less frequent conformational change observed in our 1 µs simulations. Less frequent exposure of the hydrophobic patch, would result in decreased myofilament Ca$^{2+}$ sensitivity but higher relative Ca$^{2+}$ affinity.
None of the identified functionally divergent residues that represent variation between TnC1 paralogs across teleost species were linked to structural changes in this study (Figure 6-8). However, many of these residues displayed changes in SLRF between paralogs far greater than those seen in Helix C. Highly coevolving residues are often found in flexible regions and facilitate structural transitions by forming and disrupting interactions cooperatively (68). To fully understand the dynamics between the open and closed state of TnC, TnI is necessary for stabilization of the open state, This is one factor that results in variation of Ca$^{2+}$ affinity measurements in the presence of the other constituent members of the Tn complex and thin filament (69). Interaction sites that mediate important functions by binding regulatory proteins have strong evolutionary constraints on amino acid substitutions (70). The conformation of TnC increases in degree of hydrophobic patch opening upon interaction with the switch region of TnI in mammalian models (14,71). The TnI switch peptide (TnISW) will bind to and stabilize the open conformation of TnC, while simultaneously occluding the hydrophobic cleft (14). Hydrophobic interactions between N-TnC and TnISW, and interactions between water and the hydrophilic side-chains of the solvent exposed portion of the TnI switch peptide make this a much more favourable interaction. This effect of hydrophobic switch peptide binding stabilizing the Ca$^{2+}$ bound open form increases the Ca$^{2+}$ affinity of N-TnC and extends the amount of time in this open state, resulting in a longer time with bound cross-bridges of the sarcomere. Since this interaction directly influences the conformational flexibility, the subtle changes observed between ZF TnC paralogs in this study may either be intensified or minimized with the inclusion of binding partners.

6.6. Summary

In many species reduced temperature causes a decrease in myofilament Ca$^{2+}$ sensitivity, an effect that is linked to TnC1 (9,72). In cardiac function of ectothermic species a greater resilience in the face of acute and seasonal temperature change is required. The differentially expressed TnC1b protein in teleost fish may play a role in this temperature tolerance. TnC1b has higher Ca$^{2+}$ affinity at both temperatures according to ITC and PMF measurements, and is expressed 3:1 over TnC1a in the atrium of warm acclimated
zebrafish (1). This variation in TnC Ca\textsuperscript{2+} binding between paralogs are dictated by three residue substitutions that alter the structure of helices C and D, and modify the energetic landscape of the TnC conformational change. Given that higher relative Ca\textsuperscript{2+} sensitivity is expected at lower temperatures (67), higher affinity of TnC1b is unexpected. However, the residue variants in helix C are not conserved in TnC1b across teleosts and are likely to be necessary in fish living at warmer temperatures, such as ZF (77). Based on the ITC and MD simulation evidence presented here, TnC1a is more likely to transition between the closed and open states, which should create higher myofilament Ca\textsuperscript{2+} sensitivity while TnC1b has higher ITC measured Ca\textsuperscript{2+} affinity conferred by the more stable closed conformation and less frequent conformational change. We propose that temperature-dependent effect of TnC paralog substitution is influenced by differences in the favorability of the TnC conformational change, which transduces the Ca\textsuperscript{2+} binding signal to the myofilament. Minor differences between the function of paralogs must be explained further through interactions with other components of the thin filament.

6.7. Acknowledgements

This study was supported by grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada to GFT and to DPT and CIHR to GFT. GFT is a Tier I Canada Research Chair and DPT is an Alberta Innovates Health Solutions Scientist and Alberta Innovates Technology Futures Strategic Chair in (Bio)Molecular Simulations. Molecular dynamics simulations were carried out on the Westgrid & Cacul Quebec Complexes, which are under the aegis of Compute Canada. We are grateful for GROMACS tutorials by Justin Lemkul and alchemy.org.
6.8. References


6.9. Supplementary

Figure S1. RMSD as a function of time for each temperature paralog combination. These indicate that each of the simulations has diverged substantially from the starting coordinates. Plots are a running average over 5% of the total number of data points. A) 100 ns simulations of TnC+Ca$^{2+}$ preceding PMF calculations B) 200 ns simulations of the TnC+TnI complexes C) 1 µS simulations of TnC + Ca$^{2+}$ D) 1µs simulations of TnC in the absence of Ca$^{2+}$.

1A)
1D)

![Graphs showing RMSD (Å) over time (ns) for different conditions.](image)

- cTnC (18°C)
- cTnC (28°C)
- sSsTnC (18°C)
- sSsTnC (28°C)
Table S1a. Homology model quality indicators for representative structures from 100 ns TnC+Ca$^{2+}$ simulations

<table>
<thead>
<tr>
<th></th>
<th>cTnC (18°C)</th>
<th>cTnC (28°C)</th>
<th>ssTnC (18°C)</th>
<th>ssTnC (28°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RAMPAGE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored</td>
<td>85 (98.84%)</td>
<td>83 (96.51%)</td>
<td>84 (97.67%)</td>
<td>86 (100.00%)</td>
</tr>
<tr>
<td>Allowed</td>
<td>1 (1.16%)</td>
<td>3 (3.49%)</td>
<td>2 (2.33%)</td>
<td>0 (0.00%)</td>
</tr>
<tr>
<td>Outlier</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
</tr>
<tr>
<td><strong>PROCHECK</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bad Backbone Bonds</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Bad Backbone Angles</td>
<td>4.60%</td>
<td>5.40%</td>
<td>4.10%</td>
<td>4.60%</td>
</tr>
<tr>
<td>Bad Contacts</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Molprobity Score</td>
<td>1.62</td>
<td>1.69</td>
<td>1.54</td>
<td>1.7</td>
</tr>
<tr>
<td>QMEAN Score</td>
<td>0.627</td>
<td>0.669</td>
<td>0.767</td>
<td>0.69</td>
</tr>
<tr>
<td><strong>Whatcheck Structure Z-Score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Generation Packing</td>
<td>-1.836</td>
<td>-1.079</td>
<td>-1.237</td>
<td>-1.322</td>
</tr>
<tr>
<td>2nd Generation Packing</td>
<td>-0.89</td>
<td>-1.245</td>
<td>-1.349</td>
<td>-1.079</td>
</tr>
<tr>
<td>$\chi_1$/$\chi_2$ Rotamer Normality</td>
<td>-1.45</td>
<td>-1.685</td>
<td>-2.472</td>
<td>-1.836</td>
</tr>
<tr>
<td>Backbone Conformation</td>
<td>0.672</td>
<td>0.93</td>
<td>1.177</td>
<td>0.662</td>
</tr>
<tr>
<td>Inside/Outside</td>
<td>0.955</td>
<td>0.966</td>
<td>0.926</td>
<td>0.999</td>
</tr>
<tr>
<td><strong>Whatcheck RMS Z-Score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond Lengths</td>
<td>0.644</td>
<td>0.635</td>
<td>0.642</td>
<td>0.662</td>
</tr>
<tr>
<td>Bond Angles</td>
<td>1.173</td>
<td>1.175</td>
<td>1.109</td>
<td>1.177</td>
</tr>
<tr>
<td>Omega Angle Restraints</td>
<td>1.471</td>
<td>1.235</td>
<td>1.422</td>
<td>1.369</td>
</tr>
<tr>
<td>Side Chain Planarity</td>
<td>1.886</td>
<td>1.732</td>
<td>1.488</td>
<td>1.789</td>
</tr>
<tr>
<td>Improper Dihedral</td>
<td>1.175</td>
<td>1.163</td>
<td>1.211</td>
<td>1.255</td>
</tr>
</tbody>
</table>
Table S1b. Homology model quality indicators for representative structures of TnC+Ca\textsuperscript{2+} in complex with Tn\textsubscript{sw}

<table>
<thead>
<tr>
<th></th>
<th>cTnC (18\textdegree)</th>
<th>cTnC (28\textdegree)</th>
<th>ssTnC (18\textdegree)</th>
<th>ssTnC (28\textdegree)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RAMPAGE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored</td>
<td>99 (99.00%)</td>
<td>96 (96.00%)</td>
<td>94 (94.00%)</td>
<td>98 (98.00%)</td>
</tr>
<tr>
<td>Allowed</td>
<td>1 (1.00%)</td>
<td>4 (4.00%)</td>
<td>6 (6.00%)</td>
<td>2 (2.00%)</td>
</tr>
<tr>
<td>Outlier</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
</tr>
<tr>
<td><strong>PROCHECK</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bad Backbone Bonds</td>
<td>0.00%</td>
<td>0.20%</td>
<td>0.00 %</td>
<td>0.00%</td>
</tr>
<tr>
<td>Bad Backbone Angles</td>
<td>4.30%</td>
<td>4.30%</td>
<td>5.00%</td>
<td>6.40%</td>
</tr>
<tr>
<td>Bad Contacts</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td><strong>Molprobity Score</strong></td>
<td>1.22</td>
<td>1.59</td>
<td>1.93</td>
<td>2.25</td>
</tr>
<tr>
<td><strong>QMEAN Score</strong></td>
<td>0.563</td>
<td>0.512</td>
<td>0.515</td>
<td>0.463</td>
</tr>
<tr>
<td><strong>Whatcheck Structure Z-Score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Generation Packing</td>
<td>-1.027</td>
<td>-0.607</td>
<td>-1.175</td>
<td>-1.329</td>
</tr>
<tr>
<td>2nd Generation Packing</td>
<td>0.614</td>
<td>0.548</td>
<td>-1.175</td>
<td>-0.09</td>
</tr>
<tr>
<td>$\chi_1/\chi_2$ Rotamer Normality</td>
<td>-1.924</td>
<td>1.122</td>
<td>-0.118</td>
<td>-2.428</td>
</tr>
<tr>
<td>Backbone Conformation</td>
<td>0.856</td>
<td>0.959</td>
<td>-1.724</td>
<td>1.234</td>
</tr>
<tr>
<td>Inside/Outside</td>
<td>0.988</td>
<td>0.959</td>
<td>0.928</td>
<td>1.077</td>
</tr>
<tr>
<td><strong>Whatcheck RMS Z-Score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond Lengths</td>
<td>0.505</td>
<td>0.600</td>
<td>0.485</td>
<td>0.547</td>
</tr>
<tr>
<td>Bond Angles</td>
<td>-1.027</td>
<td>1.157</td>
<td>1.216</td>
<td>1.223</td>
</tr>
<tr>
<td>Omega Angle Restraints</td>
<td>1.297</td>
<td>1.419</td>
<td>1.608</td>
<td>1.638</td>
</tr>
<tr>
<td>Side Chain Planarity</td>
<td>2.012</td>
<td>1.767</td>
<td>1.972</td>
<td>2.083</td>
</tr>
<tr>
<td>Improper Dihedral</td>
<td>1.305</td>
<td>1.19</td>
<td>1.447</td>
<td>1.281</td>
</tr>
</tbody>
</table>
Table S1c. Homology model quality indicators for representative structures of Ca\textsuperscript{2+} -free TnC from 1μs simulations

<table>
<thead>
<tr>
<th></th>
<th>cTnC (18\textdegree)</th>
<th>cTnC (28\textdegree)</th>
<th>ssTnC (18\textdegree)</th>
<th>ssTnC (28\textdegree)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RAMPAGE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored</td>
<td>85 (98.80%)</td>
<td>83 (96.50%)</td>
<td>84 (97.70%)</td>
<td>84 (97.70%)</td>
</tr>
<tr>
<td>Allowed</td>
<td>1 (1.20%)</td>
<td>3 (3.50%)</td>
<td>2 (23.00%)</td>
<td>2 (23.00%)</td>
</tr>
<tr>
<td>Outlier</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
</tr>
<tr>
<td><strong>PROCHECK</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bad Backbone Bonds</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Bad Backbone Angles</td>
<td>5.60%</td>
<td>4.60%</td>
<td>4.30%</td>
<td>4.60%</td>
</tr>
<tr>
<td>Bad Contacts</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Molprobity Score</td>
<td>0.83</td>
<td>0.82</td>
<td>0.57</td>
<td>1.37</td>
</tr>
<tr>
<td>QMEAN Score</td>
<td>0.842</td>
<td>0.605</td>
<td>0.794</td>
<td>0.72</td>
</tr>
<tr>
<td><strong>Whatcheck Structure Z-Score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Generation Packing</td>
<td>-0.916</td>
<td>-0.814</td>
<td>-0.855</td>
<td>-1.237</td>
</tr>
<tr>
<td>2nd Generation Packing</td>
<td>-0.658</td>
<td>-0.518</td>
<td>-0.365</td>
<td>-1.074</td>
</tr>
<tr>
<td>$\chi_1$/\chi_2 \text{ Rotamer Normality}</td>
<td>-2.497</td>
<td>-2.265</td>
<td>-1.093</td>
<td>-0.818</td>
</tr>
<tr>
<td>Backbone Conformation</td>
<td>1.771</td>
<td>1.23</td>
<td>1.23</td>
<td>0.934</td>
</tr>
<tr>
<td>Inside/Outside</td>
<td>0.961</td>
<td>0.984</td>
<td>0.982</td>
<td>0.955</td>
</tr>
<tr>
<td><strong>Whatcheck RMS Z-Score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond Lengths</td>
<td>0.513</td>
<td>0.553</td>
<td>0.551</td>
<td>0.501</td>
</tr>
<tr>
<td>Bond Angles</td>
<td>1.156</td>
<td>1.241</td>
<td>1.17</td>
<td>1.232</td>
</tr>
<tr>
<td>Omega Angle Restraints</td>
<td>1.428</td>
<td>1.377</td>
<td>1.512</td>
<td>1.38</td>
</tr>
<tr>
<td>Side Chain Planarity</td>
<td>1.524</td>
<td>1.69</td>
<td>1.774</td>
<td>1.496</td>
</tr>
<tr>
<td>Improper Dihedral</td>
<td>1.177</td>
<td>1.206</td>
<td>1.16</td>
<td>1.224</td>
</tr>
</tbody>
</table>
Table S1d. Homology model quality indicators for representative structures of TnC+Ca\textsuperscript{2+} from 1µs simulations

<table>
<thead>
<tr>
<th></th>
<th>cTnC (18°C)</th>
<th>cTnC (28°C)</th>
<th>ssTnC (18°C)</th>
<th>ssTnC (28°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RAMPAGE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored</td>
<td>84 (97.70%)</td>
<td>85 (98.80%)</td>
<td>81 (94.20%)</td>
<td>85 (98.80%)</td>
</tr>
<tr>
<td>Allowed</td>
<td>2 (23.00%)</td>
<td>1 (1.20%)</td>
<td>5 (5.80%)</td>
<td>1 (1.20%)</td>
</tr>
<tr>
<td>Outlier</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
</tr>
<tr>
<td><strong>PROCHECK</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bad Backbone Bonds</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Bad Backbone Angles</td>
<td>7.10%</td>
<td>7.60%</td>
<td>6.10%</td>
<td>4.40%</td>
</tr>
<tr>
<td>Bad Contacts</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Molprobity Score</td>
<td>1.2</td>
<td>0.57</td>
<td>1.08</td>
<td>0.81</td>
</tr>
<tr>
<td><strong>QMEAN Score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Generation Packing</td>
<td>-0.875</td>
<td>-1.506</td>
<td>-2.138</td>
<td>-0.616</td>
</tr>
<tr>
<td>2nd Generation Packing</td>
<td>-0.398</td>
<td>-1.155</td>
<td>-2.061</td>
<td>-0.606</td>
</tr>
<tr>
<td>$\tilde{\chi}_1/\tilde{\chi}_2$ Rotamer Normality</td>
<td>-3.038</td>
<td>-2.781</td>
<td>-3.188</td>
<td>-2.315</td>
</tr>
<tr>
<td>Backbone Conformation</td>
<td>1.06</td>
<td>1.013</td>
<td>0.356</td>
<td>0.948</td>
</tr>
<tr>
<td>Inside/Outside</td>
<td>1.019</td>
<td>0.985</td>
<td>0.993</td>
<td>0.95</td>
</tr>
<tr>
<td><strong>Whatcheck RMS Z-Score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond Lengths</td>
<td>0.488</td>
<td>0.525</td>
<td>0.542</td>
<td>0.495</td>
</tr>
<tr>
<td>Bond Angles</td>
<td>1.265</td>
<td>1.273</td>
<td>1.209</td>
<td>1.122</td>
</tr>
<tr>
<td>Omega Angle Restraints</td>
<td>1.73</td>
<td>1.554</td>
<td>1.57</td>
<td>1.786</td>
</tr>
<tr>
<td>Side Chain Planarity</td>
<td>2.332</td>
<td>1.775</td>
<td>1.686</td>
<td>1.797</td>
</tr>
<tr>
<td>Improper Dihedral</td>
<td>1.441</td>
<td>1.31</td>
<td>1.224</td>
<td>1.258</td>
</tr>
</tbody>
</table>
Table S2.  Interhelical angle frequencies over long timescale simulations.

<table>
<thead>
<tr>
<th>Temperature</th>
<th>18°C</th>
<th>28°C</th>
<th>18°C</th>
<th>28°C</th>
<th>18°C</th>
<th>28°C</th>
<th>18°C</th>
<th>28°C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ca2+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>&lt;90</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>&lt;95</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>&lt;100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>66</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>&lt;105</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>283</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>&lt;110</td>
<td>0</td>
<td>36</td>
<td>0</td>
<td>876</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>&lt;115</td>
<td>2</td>
<td>121</td>
<td>0</td>
<td>1743</td>
<td>2</td>
<td>4</td>
<td>10</td>
<td>56</td>
</tr>
<tr>
<td>&lt;120</td>
<td>25</td>
<td>447</td>
<td>1</td>
<td>2920</td>
<td>13</td>
<td>24</td>
<td>114</td>
<td>210</td>
</tr>
<tr>
<td>&lt;125</td>
<td>205</td>
<td>2147</td>
<td>19</td>
<td>7330</td>
<td>109</td>
<td>540</td>
<td>1056</td>
<td>1189</td>
</tr>
<tr>
<td>&lt;130</td>
<td>3123</td>
<td>9532</td>
<td>216</td>
<td>12210</td>
<td>2426</td>
<td>5932</td>
<td>5761</td>
<td>5745</td>
</tr>
<tr>
<td>&gt;130</td>
<td>21666</td>
<td>15420</td>
<td>24744</td>
<td>5412</td>
<td>22456</td>
<td>18497</td>
<td>18142</td>
<td>18059</td>
</tr>
</tbody>
</table>
Chapter 7.
Functional divergence in teleost cardiac troponin paralogs guides variation in the interaction of TnI switch region with TnC

This chapter is largely based on the following original research article:

Genge, C.E., Stevens, C.M, Davidson, W.S., Singh, G., Tieleman, D.P. and Tibbits G.F. (2016) Functional divergence in teleost troponin paralogs guides variation in the interaction of the TnI switch region with TnC. Genome Biology and Evolution

My contributions to this study included experimental design, evolutionary analysis and data interpretation. I was primarily responsible for writing the manuscript under the advisement of the co-authors.
7.1. Abstract

Gene duplication results in extra copies of genes that must co-evolve with their interacting partners in multimeric protein complexes. The cardiac troponin (Tn) complex, containing TnC, TnI and TnT, forms a distinct functional unit critical for the regulation of cardiac muscle contraction. In teleost fish, the function of the Tn complex is modified by the consequences of differential expression of paralogs in response to environmental thermal challenges. In this paper, we focus on the interaction between TnI and TnC, coded for by genes that have independent evolutionary origins, but the co-operation of their protein products has necessitated coevolution. In this study we characterize functional divergence of TnC and TnI paralogs, specifically the inter-related roles of regulatory sub-functionalization and structural sub-functionalization.

We determined that differential paralog transcript expression in response to temperature acclimation results in three combinations of TnC and TnI in the zebrafish heart: TnC1a/TnI1.1; TnC1b/TnI1.1 and TnC1a/TnI1.5. Phylogenetic analysis of these highly conserved proteins identified functionally divergent residues in TnI and TnC. The structural and functional effect of these Tn combinations was modeled with molecular dynamics simulation to link divergent sites to changes in interaction strength. Functional divergence in TnI and TnC were not limited to the residues involved with TnC/TnI switch interaction, which emphasizes the complex nature of Tn function. Patterns in domain-specific divergent selection and interaction energies suggest that substitutions in the TnI switch region are crucial to modifying TnI/TnC function to maintain cardiac contraction with temperature changes. This integrative approach introduces Tn as a model of functional divergence that guides the co-evolution of interacting proteins.

7.2. Introduction

Cardiac function in the ectothermic fish heart has many species-specific differences in tolerance and plasticity, which are correlated with habitat conditions and thermal preferences (1). In ectotherms, cardiac remodeling may be influenced by changes
in the susceptibility of the myofilament contractile apparatus to cytosolic calcium (Ca\textsuperscript{2+}) fluctuations. Decreases in environmental temperature reduce the Ca\textsuperscript{2+} sensitivity of the cardiac contractile unit in both mammals (2) and fish (3) despite the critical role of myofilament Ca\textsuperscript{2+} sensitivity in regulating cardiac contractility. Ca\textsuperscript{2+} handling is the primary basis for the regulation of cardiac contraction in all vertebrate species, and is determined by Ca\textsuperscript{2+} delivery to and removal from the troponin complex (Tn). Cardiac troponin (cTn) is made up of three proteins: Ca\textsuperscript{2+} binding troponin C (TnC), inhibitory troponin I (TnI), and tropomyosin-binding troponin T (TnT). The cTn complex regulates actomyosin cross-bridge interactions, in a Ca\textsuperscript{2+}-dependent manner, mediating the transition between contraction (systole) and relaxation (diastole) in the heart (4).

The importance of the role of TnC is demonstrated by its high degree of conservation through hundreds of millions of years of evolution across phylogenetically diverse groups of organisms (5). Subtle variations in TnC amino acid sequence are found in vertebrates as interspecies orthologs and intraspecies paralogs. The presence and differential expression of these paralogs implies a distinct function for each paralog. TnC2 (tnnC2, fsTnC in mammals) and TnC1 (tnnC1, cTnC in mammals) are sub-functionalized duplicates of the ancestral TnC gene. Each TnC paralog has a distinct expression pattern as well as variation in Ca\textsuperscript{2+} sensitivity (5), which has resulted in the traditional Tn protein nomenclature in mammals reflecting tissue distribution (Table 7-1). However, in sequenced fish genomes, a third paralog is present, resulting in two genes orthologous to mammalian TnC1: TnC1a (tnnC1a, cTnC, zgc:103465) and fish-specific TnC1b (tnnC1b, ssTnC, zgc:86932). TnC1b possesses a distinct expression pattern from TnC1a and TnC2 (6), and can ‘rescue’ the embryonic hearts of TnC1a knockout zebrafish embryos (7), demonstrating an overlap in the basic function of these gene products.
Table 7-1  **Nomenclature for Tn paralogs.**

Accession numbers are listed in Supplementary Table 1. Teleost nomenclature (represented by zebrafish in Ensembl) refers to terminology used to discuss teleost paralogous groups in this manuscript. The variation in naming of TnI paralogs in zebrafish is reflective of TnC and TnI2 genes that appear to be products of tandem gene duplications based on chromosomal localization, whereas TnI1 genes are all located on separate chromosomes.

<table>
<thead>
<tr>
<th>Protein Product</th>
<th>Gene Name</th>
<th>Nomenclature</th>
<th>Gene Name</th>
<th>Nomenclature</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TnC1</strong></td>
<td>TnC1</td>
<td>tnn1</td>
<td>cTnC</td>
<td>tnn1a, cTnC, s, zgc:103465, tnn1b, ssTnC, zgc:86932</td>
</tr>
<tr>
<td><strong>TnC2</strong></td>
<td>TnC2</td>
<td>tnn2</td>
<td>fTnC</td>
<td>tnn2</td>
</tr>
<tr>
<td><strong>TnI1</strong></td>
<td>TnI1</td>
<td>tnn1</td>
<td>ssTnI</td>
<td>tnn1a, zgc:103458, tnn1d, zgc:92233, tnn1c, zgc:153662, tnn1a1, zgc:86895</td>
</tr>
<tr>
<td><strong>TnI2</strong></td>
<td>TnI2</td>
<td>tnn2</td>
<td>fTnI</td>
<td>tnn2a1, zgc:103650, tnn2a2, zgc:56294, tnn2a3, zgc:92053, tnn2a4, zgc:86800, tnn2b1, zgc:110715, tnn2b2, zgc:92191</td>
</tr>
<tr>
<td><strong>TnI3</strong></td>
<td>TnI3</td>
<td>tnn3</td>
<td>cTnI</td>
<td>N/A</td>
</tr>
</tbody>
</table>

The functions of TnC paralogs differ directly through modifications of the affinity for Ca\(^{2+}\) at site II or Ca\(^{2+}\) sensitivity that includes complex conformational changes that enable muscle contraction. A number of Ca\(^{2+}\) binding proteins show variation in sensitivity between species-specific orthologs in order to function optimally over the native temperature range for a given species (8-10). While site II is conserved across TnC1 paralogs, sequence variation of TnC paralogs can manifest as variation in TnC sensitivity due to changes in protein stability or flexibility. TnC function can also differ indirectly...
through effects produced by substitution of paralogous proteins that form the Tn complex. Some paralog combinations increase the likelihood of a Ca$^{2+}$ binding event that triggers cardiac contraction. Variation in the Ca$^{2+}$ sensitivity of TnC alone does not explain the variation in the teleost contractile element relative to mammals (3,11) but the strength of interaction between multiple binding partners influences Ca$^{2+}$ sensitivity. TnI, in particular, is known to directly modulate the Ca$^{2+}$ sensitivity of TnC and thus affect the contractility of the cardiac sarcomere (12). In mammals, three tissue-specific TnI paralogs are found: TnI1 in slow-skeletal muscle (tnnI1, ssTnI), TnI2 in fast-skeletal muscle (tnnI2, fsTnI) and TnI3 in cardiac muscle (tnnI3, cTnI) (reviewed in (13)). TnI3, with a unique N-terminal extension, is specific to Sarcopterygians. Teleosts have up to seven homologs of mammalian TnI1, each with distinct tissue-specific expression profiles for transcripts and protein products (14). The variation in sequence between these teleost TnI paralogs may influence the Ca$^{2+}$ sensitivity of the contractile element to a greater extent than variation in TnC alone. An example of the consequence of sequence variation in mammals is the interaction between the TnI switch arm and the N-terminal domain of TnC (15). In particular, four amino acid substitutions in human TnI3 are responsible for increased contractility and slowed relaxation compared to TnI1 (16). Phylogenetic data suggest that the switch arm is a site of rapid divergence that provides the genetic variation that differentiates the mammalian TnI3 from TnI1 paralogs present in all other vertebrates (17).

These generalizations, however, have not accounted for the presence of multiple paralogs that result from gene duplication events in the teleost lineage. The redundancy created by duplication usually leads to the loss of copies unless a new function is found. Gene duplication may lead to functional divergence of proteins, resulting in both structural changes (18), and variation in gene expression patterns (19). When both copies undergo mutations, sub-functionalization of the ancestral function can rapidly occur and fixate in the genome if those mutations prove complementary (20,21). The division of ancestral function may be followed by neo-functionalization over time (22,23) through novel regulatory patterns and/or novel structural specificity (24). The presence of multiple paralogs is complicated further by coevolution of interacting proteins. Studies of interacting proteins have found correlated evolution of the sequences of binding partners as a result of compensating mutations to maintain specificity (25,26). Complementary changes in interacting proteins, where selection pressure on one will affect the other and vice versa,
may be both intermolecular and intramolecular. Typically, many studies have looked at complementary changes on the intramolecular level, where changes within a protein are important to proper functioning (reviewed in (27)). The globin family of proteins has provided an example of both conservation of interactions between helices to maintain the tertiary structure via common packing patterns as well as functional divergence in sites critical to intermolecular interactions in the tetrameric structure of hemoglobin (28,29). However studies using cytochrome oxidase show that structural stability is not the only factor involved with constrained co-evolutionary patterns between contact points of interacting subunits (30). The regulation of gene expression is critical for maintaining the correct stoichiometric balance of multimeric proteins (31) and small changes in expression level accumulating over time can lead to both qualitative and quantitative sub-functionalization (32). These iconic examples point to the importance of understanding both structural and regulatory divergence in co-evolution of interfacing proteins.

The Tn complex constitutes an unexplored co-evolutionary model with three crucial interactions between proteins that are necessary for contractile function. Based on protein sequence similarity, TnT and TnI have a common evolutionary origin (33) while TnC has an evolutionary origin in common with other EF-hand Ca\textsuperscript{2+} binding proteins (e.g. calmodulin) (5,34). Despite the independent evolutionary origins of TnC and TnI, the functional interactions that must occur in the Tn complex may have led to correlated evolution in specific interacting domains. Structural and functional requirements constrain the sequences of paralogs in highly expressed proteins (35). Conservative substitutions in TnC sequences alone may not be able to accommodate the variation in Ca\textsuperscript{2+} sensitivity required for disparate environmental niches of vertebrates. Sequence variations appear to have a less pronounced functional impact in TnI than in TnC (17). Because TnI and TnC are intrinsically linked in function, a change in one provides a selective force for a complementary mutation in the other. In particular, the interaction between TnC and TnI switch region can modify the Ca\textsuperscript{2+} sensitivity of TnC. The TnI switch/inhibitory region, defined either as residues 148-164 (36), or as the shorter region, residues 147-163 (PDB: 1MXL) in mammalian TnI3 binds with the hydrophobic patch of TnC between helices A and B following Ca\textsuperscript{2+} binding to the regulatory site of TnC, an interaction that stabilizes the open conformation (15). The inhibitory region (IR) of TnI (residues 137-147) interacts with helix C of TnC and the unstructured region between the D helix of the N-terminal lobe and
the E-helix of the C terminal lobe (D-E linker). The domains of TnI that interact directly with TnC will be under greater selective pressure to be conserved than non-interacting regions to maintain the appropriate interface, despite that as a whole the sequence of TnI is far less conserved than TnC across vertebrate phylogeny. Domain-specific interactions suggest that similar rates of evolution may not be observed across the entire sequence of these genes, but the rates of evolution for specific domains may be linked. However, common methods for determining evolutionary selection pressures are often limited by a lack of biophysical structural context (27) thus neglecting an important facet of structural sub-functionalization. In this study we have identified a series of amino acid sequence changes between paralogs in teleost TnI and TnC that are indicative of functional divergence. Transcriptional expression profiles were used to predict the paralog combinations of TnC/TnI in the heart of zebrafish, which were modeled to show the structural and functional effect of these substitutions. While limited structural variation was demonstrated in the divergent regions, differences in calculated interaction energies suggest that residues in the switch region that distinguish each TnI paralog have a pronounced effect on TnC/TnI function. This synergistic and mutually corroborative approach to co-evolution patterns in paralogs that form the Tn complex provides a useful paradigm that can be extended to a broad range of questions regarding diverse interacting proteins.
7.3. Methods

7.3.1. Collection of available TnI and TnC sequences

All available full-length sequences for Actinopterygii (ray-finned fishes) were collected from the NCBI (National Center for Biotechnology Information) GenBank and Ensembl (http://www.ensembl.org/index.html) (Table S1). For species with full genome information not available on Ensembl zebrafish, TnI and TnC transcript sequences were used for BLASTN searches of NCBI databases (http://www.ncbi.nlm.nih.gov) with an E value <10^-5 and default parameters. Additional salmonid sequences were found with the same zebrafish transcript sequences were compared by BLAST searches against the Atlantic salmon genome sequencing project, publicly available at ASalBase.
Salmonid and gar Tn sequences were then used for further BLASTN and BLASTP searches within the NCBI database. All sequences were translated to predicted protein products using ExPasy translator. Only full-length protein sequences were used for analyses and redundant proteins were eliminated.

7.3.2. Phylogenetic analysis

Multiple amino acid sequence alignments were performed using MUSCLE (Multiple Sequence Comparison by Log-Expectation; (38)) via MEGA6 (Molecular Evolutionary Genetics Analysis 6.0, (39)). Alignment gaps in variable regions were removed, truncating the TnI sequence. The evolutionary histories for these amino acid sequences were inferred using the maximum likelihood algorithm (MEGA6). For amino acid trees, maximum likelihood analyses were based on the JTT-matrix model (40), the optimal model of evolution built on the Akaike Information Criterion – AIC (41). Bootstrap consensus trees were inferred from 800 replicates.

Trees were generated for full-length nucleotide sequences with all available sequences from Actinopterygii and representative members of the more basal lobe-finned fish or Sarcopterygii (e.g. coelacanths) for further analysis (TnC fish only, model K2 +G, 142 sequences, 477 sites; TnI fish only - model GTR +G +I, 89 sequences, 359 sites.). Domain-specific trees were examined for interacting components of TnI/TnC using a truncated form of the sequences (TnC N-terminal end fish only – K2 +G, 89 sequences, 260 sites; TnI inhibitory region/switch domain fish only – GTR+G, 128 sequences, 108 sites). Sequence information from the Petromyzontida (e.g., lampreys) was used as an ancestral state to root the trees. Multiple topologies were compared to ensure no species bias or long-branch effects biased analyses. All optimal models of evolution were selected based on the AIC.
7.3.3. Zebrafish acclimation

All protocols dealing with zebrafish were approved by the SFU University Animal Care Committee and conformed to the Canadian Council on Animal Care (CCAC) guidelines. Zebrafish acclimation was performed as described in (6), with final temperature held for three weeks before sampling.

7.3.4. Tissue-specific transcriptional expression

Quantitative mRNA expression profiles were compared between all TnI1 paralogs with respect to acclimation temperature following the methodology described in (6). Specific primers (Table 7-2) for each gene were designed using a combination of IDT Oligo Properties calculator (https://www.idtdna.com/calc/analyzer), Primer3 (42) and Oligonucleotide Properties Calculator (43) to amplify a single product (using zebrafish published sequences – Table S2). Analysis was performed according to the \( \Delta \Delta Ct \) method using the geometric mean of housekeeping genes to normalize the data (44). Housekeeping genes included \( \beta \)-actin (45) as well as two primer sets for elongation factor-1 alpha to check for RNA stability (46). All reference genes were confirmed by both dissociation curves and the stability of Ct values across conditions and samples.

7.3.5. Molecular Dynamics (MD)

The initial models for the TnC/TnI constructs were generated using the Swiss-model workspace (47). These models used the NMR-derived human cardiac TnC structure (PDB:1MXL) (15) as a template. The zebrafish TnC1a and TnC1b sequences, and the switch regions of TnI (residue 116-132 in zebrafish TnI1.1) were used as target sequences. The TnC/TnI complex pdb files for each TnI/TnC paralog combination were assembled through superimposition with the 1MXL structure. The resulting models were simulated with GROMACS 4.6.2 (48), using the AMBER99-sb-ILDN (49) force field. The simulation system was defined as a periodic cubic box with 1 nm spacing between the edge of the box and the nearest protein atom. The system was solvated using the TIP3P
water model (50), and made neutral by replacing randomly-selected water molecules with K+ ions. An additional pair of K+ and Cl- ions was also added to each simulation system. This was followed by steepest descent energy minimization to a tolerance of 10 kJ mol⁻¹ nm⁻¹ and conjugate gradient energy minimization for 10,000 steps. The minimized system was restrained with 1,000 kJ mol⁻¹ nm⁻¹ absolute position restraints on all of the non-solvent atoms and equilibrated for 1 ns.

These restrained simulations were held at 28°C using V-rescale temperature coupling (51) and isotropic Berendsen pressure coupling (52) with a $\tau_T$ of 0.1 and $\tau_P$ of 1.0, respectively. Interactions were calculated using PME electrostatics (53) and the Verlet cut-off scheme (54). Bond lengths were constrained using the LINCS algorithm (55).

Production simulations were done in 5 replicated unrestrained 200 ns NPT simulations; other parameters were carried forward from the position-restrained simulations. The final models were produced by clustering with the Daura algorithm (56) over the backbone and C-beta atoms of each structure across the five trajectory replicates of each paralog combination; the middle structure of each of the five largest clusters from each TnI/TnC paralog combination were used for further analysis. Protein structure superimpositions were performed with VMD (57), and structural quality assessments were carried out using RAMPAGE (58), QMEAN (59), WHATCHECK (60), PROCHECK (61) and MOLPROBITY (62).

The interface analysis between TnI and TnC was performed using g_mindist (48) over the final 20 ns for each trajectory, and minimum distances were averaged over the five replicated simulations. The interface surface area was calculated with UCSF Chimera (63) and Intersurf (64). The surface areas were calculated for representative structures from each of the five most common clusters and the resulting average surface area was weighted according to the number of structures in each cluster. The estimation of binding free energy was calculated with g_mmpbsa (65) using snapshots 1 ns apart from the final 20 ns of each simulation. The MM/PBSA calculations used the non-linear Poisson-Boltzmann equation and calculations were performed at 28°C, with a solvent dielectric constant of 80, and probe radius of 1.4 Å.
7.3.6. dN/dS ratios for selective pressures

With mutual evolution, similar selective pressures/evolutionary history patterns are expected for interacting domains since they will be under complementary constraints. To detect evidence of either positive or purifying selection, the number of replacement mutations per replacement sites (nonsynonymous dN) was compared to the number of silent mutations per silent site (synonymous dS) represented as $\omega$ using the codon-based maximum likelihood (codeML) package as part of PAML 4.7 (66). PAML's Branch model (BM2), allowing the starting $\omega$ ratio to vary among lineages but not within sites within the protein, was used to detect positive selection on particular lineages (66). BM2 was compared to the null as nested models using critical values of the Chi square distribution using the likelihood ratio test (LRT) statistic, and degrees of freedom as the difference in the number of parameters were estimated with each model. To identify codon sites experiencing divergent selection between TnC paralogs as well as TnI paralogs Clade Model C (CM3) (67) was compared to the M2a_rel null model (68). This model partitions the sites among three site classes, with $\omega$ estimated for each site class. This test was only run on fish-specific trees since incorporating a broader scope of phylogeny would become too divergent on the nucleotide scale for accuracy. The fit of each model was compared to the null model by the Log Likelihood values (LnL). As in BM2, the LRT was used to compare the null model with CM3, which differs only by the assumption of divergent selective pressures at one class of sites following a duplication event. Significance in the LRT indicates the presence of sites evolving under significantly different selection pressures between the two clades. The Bayes Empirical Bayes (BEB) test was then used to determine to which class an individual codon site is likely to belong (69,70). To avoid detection of local peaks, all analyses were performed with a variety of starting $\omega$ values (0, 0.4, 1). A variety of phylogenetic trees were used to avoid incomplete taxon sampling or species bias (71-73).

7.3.7. Functional divergence of amino acids

Functional divergence comparing the predicted proteins of genes in each clade of interest was calculated using the statistical framework provided in the GU99 method of
DIVERGE 3.0 (74). Type I and type II divergence were compared between clusters of TnC (1a vs 1b) and Tnl (1.1 vs 1.5). Coefficients of $\theta_I$ and $\theta_{II}$ significantly >0 demonstrate divergence between clusters. For type I functional divergence, to reject the null hypothesis ($\delta \theta = 0$), we compared two times the standard error of theta and a LRT (75). For type II functional divergence, pairs with $\theta_{II}$ values greater than 0 after subtracting two times the standard error were considered significant.

### 7.4. Results

#### 7.4.1. Phylogenetic analyses

There are three paralogs of TnC across teleosts (Figure 7-1). TnC1a and TnC1b are orthologous to mammalian TnC1 (cardiac). They appear to be the product of a tandem duplication prior to the teleost-specific whole genome duplication (Ts3R). The evidence for this comes from both paralogs being in the genome of gar (*Lepisosteus oculatus*), a pre-Ts3R outgroup to teleostei, and a single TnC1 being present in the genomes of any higher vertebrates classified as part of Sarcopterygia.

In zebrafish we found six potential Tnl1 paralogs. Mammals have three Tnl paralogs – Tnl1 (slow skeletal), Tnl2 (fast skeletal) and Tnl3 (cardiac) (Figure 7-2); Tnl3 exists as an independent clade specific to endotherms. In fish there appear to be three sub-clades associated with Tnl1. In all fish taxa with sequence data available, at least one paralog exists in each of these Tnl1-like sub-clades, while some species have two paralogs for each of these sub-clades. Mammalian Tnl1 groups with predominantly Tnl1.1/1.2, the paralog most commonly found in the literature as ‘fish ssTnl’ or Tnl1b. The second sub-clade represents Tnl1.3/1.4. Paralogs from this clade are often listed on Ensembl as Tnl1c or Tnl1d. The third sub-clade is Tnl1.5/1.6, seen on Ensembl as Tnl1al.

In all teleost species, for which chromosomal information is available (zebrafish, medaka, tetraodon, stickleback), all three TnC genes are located on one chromosome whereas the two TnC genes in mammals are on separate chromosomes. The three TnC
genes are all located on chromosome 23 in zebrafish, supporting a tandem duplication with a possible translocation of TnC1a further away from TnC1b and TnC2. All paralogous TnI1 genes are on separate chromosomes in fish genomes with chromosomal data available and are located in close proximity to a TnT2 paralog, which suggests possible Ts3R products. In spotted gar, TnC1a and TnC1b both are located on different chromosomes, as well as on separate chromosomes from TnI1 and TnT2. This is also seen in representative higher vertebrates, where TnC is located on a separate chromosome while TnI3 and TnT2 appear to be linked spatially.
Figure 7-2  TnC maximum likelihood tree.
The evolutionary history was inferred using the maximum likelihood method based on the JTT model for the MUSCLE amino acid alignment of full length TnC (159 sites in final dataset). The bootstrap consensus tree was inferred from 800 replicates with the percentage of replicate trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The tree is rooted with lamprey TnC2. The analysis involved 142 amino acid sequences of representative vertebrate TnC sequences compiled from NCBI and Ensembl (Supplementary Data - Table 1). Evolutionary analyses were conducted in MEGA6 (39).
Figure 7-3  TnI maximum likelihood tree.
The evolutionary history was inferred using the maximum likelihood method based on the JTT model for the MUSCLE amino acid alignment of TnI full length sequence after the removal of gaps (121 sites in final dataset). The bootstrap consensus tree was inferred from 800 replicates with the percentage of replicate trees in which the associated taxa clustered together is shown next to the branches. The tree was drawn to scale, with branch lengths indicating the number of substitutions per site. The tree is rooted with lamprey TnI2. The analysis involved 128 amino acid sequences of representative vertebrate TnI sequences (Supplementary Data - Table 1). Evolutionary analyses were conducted in MEGA6 (39).
7.4.2. Transcript expression patterns

TnI1.1 was the predominant paralog expressed in both the zebrafish atrium and ventricle (Figure 7-3). With acclimation to cold temperature (18°C), transcript expression was significantly up-regulated in both chambers. The second highest expressed paralog in the ventricle is TnI1.5. With acclimation to warmer temperatures (28°C), TnI1.5 transcript expression was significantly increased in the ventricle only. Negligible amounts of TnI1.3, 1.4 and 1.6 transcripts were found in zebrafish cardiac tissue. These paralogs may be more important in skeletal muscle (data not shown in zebrafish; trout - (14)) and were not considered further in this study.

Figure 7-4  Relative expression levels of TnI paralogs in zebrafish hearts.
Quantitative real time PCR was used to determine tissue specific differences in relative mRNA levels of TnI paralogs in adult zebrafish at 28°C and 18°C. Values are expressed as mean (n=8) of mRNA levels normalized to the geometric mean of β-actin and Ef1α by the ΔΔCt method. Vertical error bars represent standard error of the mean. * represents p<0.05 (Student’s t-test). Note two orders of magnitude in scale bars on the ordinate for atrial TnI 1.1 vs. 1.5.
7.4.3. Equilibrium Molecular Dynamics and homology models

After 200 ns of simulation the effects of the amino acid sequence substitutions are visible in the deviation of the structures from the starting frames of each simulation (Figure S3). A representative structure was selected from the middle conformation of the largest cluster from each Tnl/TnC paralog combination. The quality indicators for the representative structures are listed in Table S5. Variation was observed between representative structures from the five largest clusters; e.g. the backbone RMSD was 2.2 Å for TnC1a with Tnl1.1, 1.9 Å for TnC1a with Tnl1.5 and 1.8 Å for TnC1b with Tnl1.1 (Figure 7-4).

The orientation and position of each of the Tnl switch peptides are slightly different in each TnC/Tnl combination (Figure 7-4) although many of the contact residues remain involved in the interaction (Figure 7-6). These demonstrate that while the regions of the protein that make contact are similar, the roles of specific interacting residues are different in each paralog combination. The combinations that include Tnl1.1 make fewer contacts overall, suggesting more specific interactions from both TnC paralogs than the TnC1a/Tnl1.5 pairing.

The interface surface areas for each of the paralog combinations are 1,530 Å², 1,561 Å² and 1,523 Å² for TnC1a/Tnl1.1, TnC1a/Tnl1.5 and TnC1b/Tnl1.1, respectively. This suggests that it is the nature of the contacts that mediates the differences between the paralogous protein interactions. The interaction surface is not limited to residues that vary between paralogs, with the exception of residues 87 and 88 of TnC, the remainder of the interfacing sites are identical in both TnC1a and TnC1b. Similarly in Tnl, the divergent sites 116, 117, 121, 123, 125 and 132 (equivalent to zebrafish Tnl1.1) are among the most common contacts in the Tnl – TnC interface.
Figure 7-5  Models and MM/PBSA interaction energies of the TnC/ TnI paralog combinations.

Homology models based on TnC/TnI switch NMR structure Pdb: 1MXL (N-terminal TnC (residues 1-89) and TnI switch (residue 116-133 in zebrafish TnI (96)). These models were generated through 200 ns Molecular Dynamics simulations to predict the protein-protein interaction surface between possible TnC/TnI paralog combinations. A) The representative structures for the 5 largest clusters superimposed. The position and orientation of the TnI is variable between replicates and across TnC/TnI combinations. There is little variation between TnC structures in the representative structures, as is the N-terminal portion of the TnI switch while the C-terminal portion of the TnI switch is variable. The greatest variation between the different clusters is observed in the TnC1a/TnI1.5 combination, which suggests a weaker interaction between these two paralogs. B) MM/PBSA Binding Energy of TnC/TnI Switch peptide combinations over the final 20 ns of each simulation, averaged over 5 replicates. The values for TnC1a and TnC1b with TnI1.1 are comparable while the TnC1a/TnI-1.5 interaction is approximately half as favourable.
7.4.4. Selection pressures

$\omega$, obtained from the codeML package in PAML 7.7, provided a measure of natural selection on the protein. Values of $\omega<1$, $\omega=1$ and $\omega>1$ indicate negative purifying selection, neutral evolution and positive selection, respectively. When $\omega$ is calculated across all sites and all lineages the average value is rarely $>1$, because positive selection is unlikely to affect all sites over a prolonged time. We focused on detecting selection in only some lineages or specific domains to eliminate bias. The domains of interest were the N-terminal domain of TnC and the switch region of TnI. For phylogenetic testing the switch region of TnI was extended by ten residues in both the N-terminal and C-terminal directions beyond what is used in the 1MXL-based structural models.

PAML’s branch model 2 was used to detect positive selection along particular lineages (foreground branches). For all tree topologies, BM2 fit the data better than the null based on the Log Likelihood values ($\text{LnL BM2}>\text{LnLnull}$). For both TnC and TnI, no significant LRTs were found when comparing any branch in particular because of the relatively small number of substitutions. This model was run across several different tree topologies, including full length and domain-specific trees (Figure S1). With no significant likelihood ratio test (LRT), it follows that no significant Bayes Empirical Bayes (BEB) was identified for any sites under positive selection. For this reason, no branch-site models allowing $\omega$ to vary both among sites in the protein and across branches in the tree were used (76,77).

The clade model C (67) was used to test for divergent selection between clades (TnC1a vs 1b; Tnl1.1 vs 1.5). This model (CM3) assumes unconstrained discrete distribution for three site classes with omega values of $<1$, $1$ and $>1$. Compared to M0, CM3 fit the data better for full-length TnC of all fish species. The LRT of CM3 for this tree topology was not significant and does not show any variation in the amount of divergent selection between TnC1a and TnC1b. However, when focusing the analysis on the N-terminal portion of TnC (residues 1-89 – Table 7-2), 78% of sites show a purifying
selection of 0.112 across all clades, while 12% of sites show divergent selection between clades TnC1a and TnC1b (0.023 and 0.033, respectively – Table 7-3).

CM3 was a better fit than M0 for the fish-specific full-length TnI tree topology (Table 7-3). This tree topology also exhibited a significant LRT (p < 0.01) for CM3, which indicates divergent selection-based variation between the selected sub-clades TnI1.1 and TnI1.5. All sites showed evidence of purifying selection, with 60% of sites showing strong purifying selection in all clades (\( \omega = 0.036 \)). Thirty seven percent of sites were found to be site class 3 (divergent) with less purifying selection (\( \omega = 0.19 \) in clade TnI1.1; \( \omega = 0.16 \) in clade TnI1.5). Focusing on the switch region of TnI in fish (residues 116-143 in zebrafish TnI1.1) shows that 66% of sites demonstrated strong purifying selection (\( \omega = 0.018 \)), 3% show neutral selection while 31% of the sites show divergent selection. The divergently evolving sites in the switch region of TnI1.1 were under stronger purifying selection pressure (\( \omega = 0.036 \)) compared to sites in TnI1.5 (\( \omega = 0.064 \)). BEB probabilities suggest that site 126, arginine (R) in zebrafish TnI1.1/ glycine (G) in zebrafish TnI1.5 (equivalent to amino acid 156 in mammalian cTnI), is a significantly divergent site.

### Table 7-2 Results of Clade model C (CM3) testing for divergent selection among codons between TnC paralogs

<table>
<thead>
<tr>
<th>Model</th>
<th>LnL</th>
<th>kappa</th>
<th>Site class 0 (all branches)</th>
<th>Site class 1 (all branches)</th>
<th>Site class 2 (clade 1,2 and background branches vary)</th>
<th>LRT test result</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TnC fish specific N-terminal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clade model C</td>
<td>-5814.07</td>
<td>1.78</td>
<td>( \omega_0 = 0.78 ) ( \omega_1 = 0.011 )</td>
<td>( \omega_1 = 0.01 ) ( \omega_1 = 1.00 )</td>
<td>( \omega_2 = 0.21 ) ( \omega_2 = 0.033 ) ( \omega_2 = 0.033 )</td>
<td>18.34 ( p&lt;0.001 )</td>
</tr>
<tr>
<td>Clade 1 - 1a</td>
<td>-5823.39</td>
<td>1.77</td>
<td>( \omega_0 = 0.77 ) ( \omega_1 = 0.01 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clade 2 - 1b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TnC fish specific full length</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clade model C</td>
<td>-11890.15</td>
<td>1.357</td>
<td>( \omega_0 = 0.74 ) ( \omega_1 = 0.013 )</td>
<td>( \omega_1 = 0.01 ) ( \omega_1 = 1.00 )</td>
<td>( \omega_2 = 0.25 ) ( \omega_2 = 0.039 ) ( \omega_2 = 0.042 )</td>
<td>0.973 ( p = 0.615 )</td>
</tr>
<tr>
<td>Clade 1 - 1a</td>
<td>-11890.632</td>
<td>1.347</td>
<td>( \omega_0 = 0.74 ) ( \omega_1 = 0.013 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clade 2 - 1b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 7-3  Results of Clade model C (CM3) testing for divergent selection among codons between TnI paralogs

<table>
<thead>
<tr>
<th>Model</th>
<th>LnL</th>
<th>kappa</th>
<th>Site class 0 (all branches)</th>
<th>Site class 1 (all branches)</th>
<th>Site class 2 (clade 1,2 and background branches vary)</th>
<th>LRT test result</th>
</tr>
</thead>
<tbody>
<tr>
<td>TnI fish specific switch/IR domain</td>
<td>128 sequences 108 sites</td>
<td>M2a_rel (null)</td>
<td>-4179.054</td>
<td>1.949</td>
<td>p0 = 0.66</td>
<td>p1 = 0.03</td>
</tr>
<tr>
<td>Clade model C</td>
<td>-4179.054</td>
<td>1.949</td>
<td>p0 = 0.66</td>
<td>p1 = 0.03</td>
<td>p2 = 0.31</td>
<td>15.6</td>
</tr>
<tr>
<td>Clade 1 - 1.1</td>
<td>-4179.054</td>
<td>1.949</td>
<td>p0 = 0.66</td>
<td>p1 = 0.03</td>
<td>p2 = 0.31</td>
<td>15.6</td>
</tr>
<tr>
<td>Clade 2 - 1.5</td>
<td>-4179.054</td>
<td>1.949</td>
<td>p0 = 0.66</td>
<td>p1 = 0.03</td>
<td>p2 = 0.31</td>
<td>15.6</td>
</tr>
</tbody>
</table>

#### 7.4.5. Functional divergence

Residues that are highly conserved in one paralog but variable in another are represented by Type I functional divergence. Due to the high level of conservation of TnC, no residues appear to be under Type I functional divergence when examining either full length or the N-terminal domain. In the TnI switch region, Type I functional divergence was detected in several residues (Figure 7-5: TnI switch, LRT = 10.69, critical value = 0.75, df=1, p<0.05). Amino acids that are highly conserved within each paralog sub-clade but vary between paralog groups are represented by Type II functional divergence. In TnC, residue 2 and 4 in the N-terminal helix, as well as residue 82 and 83 in the unstructured region joining helices D and E show significant evidence of Type II functional divergence between TnC1a and TnC1b (Figure 7-6).
Figure 7-6  Key sites undergoing divergent selection/functional divergence in the TnI switch/IR region (residue 106-143 in zebrafish TnI1.1).

Alignment across fish specific sequences demonstrates key sites conserved within each paralog appear to be sites of functional divergence. Note for phylogenetic testing the switch region of TnI was extended by 10 residues in both the N-terminal and C-terminal direction beyond what is used in the 1MXL-based structural models. The black boxes denote divergent selection sites as determined by Bayes Empirical Bayes (BEB) probability above 95% confidence using PAML clade model C (CM3). All sites detected were located in TnI and exhibited relaxed purifying selection. Purple sites denote coefficients of $\theta_1$ significantly >0, demonstrating Type I divergence (highly conserved in one paralog but variable in another) between clusters as determined by DIVERGE 3.0 (probability cut-off 0.76). Blue sites denote sites demonstrating Type II divergence (highly conserved within each paralog but vary between each paralog), with $\theta_1$ values greater than 0 after subtracting two times the standard error. See text for further details.

<table>
<thead>
<tr>
<th>TnI1.1</th>
<th>TnI1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZFL.1</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>SBI.1</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>MD1.1</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>FFL.1</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>TLI.1</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>ZMI.1</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>COD1.1</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>COD1.2</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>SBI.2</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>FUGU1.1</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>TLI.2</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>ZMI.2</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>GAR1a</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>RTL1.1</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>CFI.1</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>MTL1.1</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>ZFL.5</td>
<td>KMKRFNLRVRKSAEAMLGALTDGSRVSVTDFKANLK</td>
</tr>
<tr>
<td>ASL.5</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>SBI.5</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>CS1.5</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>RTL5.5</td>
<td>KFKRFLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>ASL.6</td>
<td>KGRKNIKRVKSAEAMLGALTEKLSKDFKANLK</td>
</tr>
<tr>
<td>FFL.5</td>
<td>KMKRFNLRVRKSAEAMLGALTDGSRVSVTDFKANLK</td>
</tr>
<tr>
<td>MDI.5</td>
<td>KMKRFNLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>ZFL.6</td>
<td>KMKRFNLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>ZMI.5</td>
<td>KMKRFNLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>FUGU1.5</td>
<td>KMKRFNLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
<tr>
<td>SBI.6</td>
<td>KMKRFNLRVRVSAADILKLLGSKHKEYSLDLRNLK</td>
</tr>
</tbody>
</table>
Figure 7-7  Interacting residues between TnI and TnC paralogs. Averaged minimum distances between residues for the final 20 ns of replicated simulations are plotted as a heat map, with darker red indicative of a closer interaction. The interactions between the TnC1a and TnI switch peptides are oriented such that the N-terminal portion of TnI interacts with the C-terminal regions of TnC, and the C-terminal region of TnI interacts with the N-terminal region of TnC. The C-terminal region of TnC1b/TnI1.1 makes contacts across the TnI sequence suggesting an expanded role for the D-E linker in TnC1b. There is a substantial difference between the TnI1.1 and TnI1.5 interfaces with TnC1a, the contacts between TnC1a and TnI 1.5 are greater in number suggesting a less specific interaction. The interactions between TnI1.1 and TnC1a are more specific, particularly with respect to the interactions with TnC residues 36-60 when compared with the TnC1a/TnI1.5 interaction.
7.5. Discussion

There are multiple genes in fish genomes that code for either TnC or TnI as a result of either the teleost-specific 3R (Ts3R) or specific tandem duplication events (78). In genes that evolve slowly, such as members of the Tn complex, it is difficult to detect large structural or coding sequence sub-functionalization, because they accumulate fewer deleterious mutations that do not have detrimental effects on function (79,80). However, the presence of divergent expression patterns based on tissue-specific profiles of both TnC and TnI paralogs indicate that regulatory sub-functionalization has occurred or is occurring for both genes (78,81). It is a further challenge to identify whether regulatory sub-functionalization can occur independently of structural sub-functionalization (24), especially in situations in which multiple proteins must interact to form a functional unit. We take a novel approach by combining traditional detection of selection pressure across genomic sequences and the detection of functional divergence across protein sequences (74). These predictions are corroborated by calculations of the divergent of protein structure and function through molecular dynamics simulation. These three methods are applied to analyze the effect of paralog substitution on interacting regions of a multimeric protein complex. The method described herein provides an integrative approach to the study of regulatory sub-functionalization, which determines the paralog composition of the complex and structural sub-functionalization in interacting proteins, which identifies the differences in the interaction.

7.5.1. Differential expression patterns of TnI paralogs combined with structural variation in TnI_{switch}/TnC interaction suggest functional divergence between paralogs

Differential expression profiles were detected for TnI paralogs with respect to chambers of the heart as well as in response to a chronic temperature perturbation in adult zebrafish (Figure 7-2). In mammals, TnC and TnI paralogs have been defined by their tissue expression profiles, which reveal that TnC1 (cTnC) and TnI3 (cTnI) are exclusively expressed in adult cardiac tissue. Mixed transcript expression is found for both TnC and TnI paralogs within teleost cardiac tissue here and in previous work with various teleosts.
(e.g., zebrafish (82,83); trout (14))). In zebrafish cardiac muscle the most likely combination of TnC and TnI paralogs to form the Tn complex is TnC1a/TnI1.1. In mammals, TnI1 and TnT2 are co-localized on the same chromosome and show linked regulation of developmental expression whereas the TnC gene appears to be independently regulated (84). In teleosts, all TnI1 genes are located in close proximity to TnT genes. However, the gene with the most common expression in cardiac tissue, TnC1a is localized with the most commonly expressed TnT2a gene (TnC - (6) TnT - (82)), but the TnI gene found in this region (TnI1.2/TnI1a) is not expressed in cardiac tissue or skeletal muscle of zebrafish (data not shown). Spotted gar, basal to teleosts, displays a gene arrangement pattern similar to mammals with synten of TnI1/TnT2 genes, as do representative amphibians and reptiles. In mammals, gene structural linkage is tied to regulation of expression patterns of TnI/TnT (85), which suggests of shared ancestral state. The presence of additional paralogs in teleosts has resulted in new patterns of gene linkage and also suggests that there are independent patterns of regulation to maintain the appropriate stoichiometric balance of Tn subunits. Further work will be necessary to examine whether complementary groups of paralogs have evolved complementary regulatory suites. While we can only speculate at this time on the existence of commonalities in cis-regulation that may be driving linked expression patterns between TnC/TnI paralog combinations, the presence of tissue-specific expression patterns suggests that both TnC and TnI paralogs are sub-functionalized on a regulatory level.

Identification of the expressed paralog combinations allows us to predict the functional consequences of changes in sequence and structure with variation in the paralog composition of the Tn complex. It is often unclear whether changes in regulatory control after gene duplication are combined with changes at the protein level (24,86) due to the difficulty of characterizing protein function based on genomic data. In the globin family, overall structure is conserved while changes in regulation drive diversity in function (28); but maintenance of structural stability permits variation in interactions between proteins. By using well-characterized proteins such as Tn, combined with structural data available for mammalian orthologs with high sequence identity, we are able to make robust predictions about functional divergence. There is little variation between the predicted structures of TnC paralogs (Figure 7-4). This does not discount sub-functionalization, as protein structures are often similar with as little as 40% amino acid sequence identity (29).
Our focus was on the interaction between the N-terminal lobe of TnC, and the switch/inhibitory regions of TnI, a critical point which governs both the Ca$^{2+}$ activation of muscle contraction and the longevity of the actomyosin complex. We have calculated differences in the interaction strengths of the simulation-based models of the three expression-profile derived configurations of TnC/TnI complexes (Figure 7-4). Intriguingly, both paralogs of TnC show similar strengths of interaction with TnI1.1. In cold-acclimated zebrafish, TnC1a/TnI1.1 is the combination of the paralogs with the highest expression level. This is consistent with the finding that fish require higher Ca$^{2+}$ sensitivity to maintain cardiac contractility and survival at lower temperatures (5), as a more favourable TnC/TnI interaction should result in higher Ca$^{2+}$ sensitivity.

7.5.2. Purifying selection occurs in all paralogs of TnI and TnC but domain-specific divergent selection patterns suggest TnI switch region variation distinguishes between TnI paralogs

As expected, highly-conserved (87) and highly-expressed proteins like TnC and TnI exhibit a high degree of purifying selection. With sub-functionalization, purifying selection may be relaxed (88), especially when the breadth of expression is decreased across tissues (76,89). While no variation was found in the amount of divergent selection between full length TnC1a and TnC1b (Table 7-4), divergent purifying selection was detected between the selected sub-clades TnI1.1 and TnI1.5 (Table 7-4). In domain-specific analysis of both TnC (N-terminal) and TnI (switch), divergent selection pressures are detected between paralogs but more divergence is present between TnI paralogs. Greater purifying selection is found in the switch region than the full-length TnI, due to the presence of an interacting motif. A previous study suggested that the mammalian TnI3 switch region is under relaxed purifying selection relative to TnI1 (17). This was not corroborated by our analyses; rather, we found that mammalian TnI1 was under greater purifying selection relative to TnI3 and fish-specific TnI 1.1/1.2, 1.5/1.6 (Table S5). The discrepancy between our study and the earlier work (17) is likely due to the significantly increased range of phylogeny used in this study decreased the possibility of longer branch lengths influencing the reliability of selective pressure estimates. The breadth of phylogeny between fish and mammals increases sequence divergence, making this comparison less statistically relevant and more susceptible to false positives (73). By limiting the
comparison to fish species, higher overall purifying selection was found in the TnI switch region, but greater divergence in selective pressure is found between paralogs than when the full-length sequences are considered. This divergent pattern suggests the switch region is critical in defining the variation in function between paralogs.

7.5.3. Type II functional divergent patterns in TnC increase variability in interaction with TnI switch region but do not influence interaction strength

Differences in evolutionary rates between gene clusters provide an opportunity to predict important amino acid residues which can be further verified through exploring the structure-function relationship at these sites (90). PAML measures of $dN/dS$ assume all sites evolve independently, making it important to consider functional divergence with respect to structural constraints (27). Sites that contribute to functional divergence have a greater effect on the protein structure than sites where functional divergence is not detected (28). In TnC, the sites identified when looking for Type II functional divergence (fixed amino acid differences between paralogs – Figure 7-6) represented residues that were unique between paralogs (discussed in (6)). The majority of the divergent sites are located in the N-helix of TnC. While the N-helix does not interact with the switch region of TnI, it has been postulated that this region modulates Ca$^{2+}$ sensitivity through intramolecular interactions that influence the structural stability of site II of TnC (91,92). The only TnC residues for which Type II functional divergence was detected and are likely to interact with the TnI switch region, are located in the D-E linker. The interfacing regions of TnI and TnC are similar in size across paralog combinations, and many of the contact residues are similar. The interaction between TnI and TnC is less diverse between the Helix D and D-E linker region of TnC and N-terminal portion of the switch region. There is more variability in the relative positions of the proteins toward the C-terminus, and despite the similarity in the number of contacts and interaction energies across replicated simulations, the precise interacting residues are not identical in each replicate (Figure 7-6, Table S6). This suggests that the hydrophobic interaction is, in part, non-specific which allows many possible, energetically similar binding conformations to exist. These result in similar interaction strengths between TnC paralogs and the TnI1.1 switch region despite the variability in specific contacts. Variability in the interaction between TnC and TnI
demonstrates the importance of using dynamic models rather than static structures to strengthen evolutionary models (93). Preferential expression of interchangeable paralogs that interact in subtly different ways, with each other and other proteins in the contractile element, enables a fine-tuning of the initiation of contraction. The appropriate combination of paralogous Tn proteins contributes to an optimized condition-dependent function rather than simply a higher or lower Ca\textsuperscript{2+} sensitivity.

7.5.4. Functionally divergent residues in TnI switch region are not all interaction points with TnC but modify interaction strength

It is not surprising that the residues that vary between TnI paralogs in the switch region are under the greatest divergent selection pressure, and are the greatest contributors to functional divergence through the modification of interaction strength with TnC. The TnI paralogs are divergent enough that despite the similarity in interfacing positions, number of contacts and interface surface area, the calculated strength of the interactions is unique for each of the TnI paralogs combined with TnC1a. We have detected examples of Type I and II functional divergence in TnI switch region. The divergent residues in this region of TnI (116, 117, 121, 123, 125 and 132 in zebrafish TnI1.1) make up a large portion of the interfacing surface of TnI, and are responsible for the differences in interaction energy between the TnI paralogs and TnC. The site that is equivalent to residue 132 in mammalian TnI1 and residue 164 in mammalian TnI3 was found to be under Type I functional divergence. This is a distinguishing residue between TnI1 and TnI3 in mammals (17), and varies between TnI1.1 (polar histidine) and TnI1.5 (non-polar valine) in teleosts. A non-polar residue (alanine) at this site in mammalian TnI3 has been associated with decreased function at low pH (94) and is not present in TnI3 in more basal vertebrates such as the anole. Histidine protonation at the equivalent residue in mammalian TnI1 is purported to maintain function in acidic conditions via interaction of H132 TnI1/ E19 TnC (95). The C-terminal region of the TnI switch peptide is very flexible in each of the three simulated paralog combinations, and as such the distance between H/V132 TnI and E19 of TnC varies over the course of the simulations. The range of mobility of residue 132 is not different between paralog combinations, and is similar to what is observed in the NMR-derived template model (PDB: 1MXL) (96). Protonation at H132 has
been suggested to induce a curved conformation of TnI (97), but the absence of a sustained interaction between H132/E19 may also be due to the lack of protonation in the simulation, which was carried out at pH 7.0. The transient nature of this interaction is due to the location of H/V132 residues of zebrafish TnI on a mobile unstructured region that is in close proximity to E19. Variability of both contacts and residue distances between the TnI switch and TnC may encompass both the extended conformation of TnI as well as the curved conformation. The functional divergence detected at residue 132 lends support to the significance of substitution at this site distinguishing fish paralogs and may promote faster sarcomeric relaxation and lower Ca\(^{2+}\) sensitivity (17), as has been found in mammalian TnI3.

Along with residue 164 in TnI3 (equivalent residue to 132 in TnI1), three other residues (Q157, E166 and H173) have been proposed to facilitate normal contraction and relaxation in the mammalian heart (16). While earlier alignments suggest that this residue configuration is specific to endotherms, this combination of residues does not exist in (endothermic) chicken TnI3 but Q157 does exist in (ectothermic) anole TnI3. Here we demonstrate that these same four residues distinguish TnI1.5 from TnI1.1. The equivalent to site Q157 in TnI3 is the only one of these four residues for which divergent purifying selection in CM3 was detected (Figure 7-5). This residue is an arginine in teleost TnI1.1 and a glycine in TnI1.5. This suggests that this site may be crucial in the weaker interaction of the switch region of TnI1.5 with TnC1a. This combination only appears in the ventricle of zebrafish acclimated to warmer temperatures, and could facilitate faster mechanical relaxation. While the contacts made by the divergent residues in the model are very similar in the N-terminal region of the TnI switch peptide, it is more difficult to describe confidently in the C-terminal switch region due to greater structural variation. Further functionally divergent sites such as residue 144 (equivalent to threonine 144 in mammalian TnI3) in the IR are beyond the simulation-based switch region model described in this work. Threonine 144 is critical for length-dependent activation in mammalian TnI3 (98) but this putative phosphorylation site varies across species as well as in other TnI paralogs. Both conservative substitutions as well as non-conservative substitutions are found at this site in many TnI1.5 sequences (Figure 7-5). Without information about whether there is variation in interaction and/or even phosphorylation by PKC at this equivalent site in teleosts it is difficult to interpret the functional relevance of divergence. Other functionally
divergent sites also may not represent critical sites in the interface of TnI switch region and TnC, but have roles in other protein interactions allowing for structural sub-functionalization without the complete loss of function. We suggest that paralog evolution has allowed more subtle modifications of Tn function through relaxed purifying pressure on modifying residues around fixed interacting sites.

7.6. Summary

There is strong support for the hypothesis that proteins that functionally interact co-evolve their structures (99). Many co-evolving sites within proteins are expected to contribute to physiologically meaningful interactions (100), but divergent sites in the TnI switch region do not all correspond to sites of significance in TnI/TnC interaction. This work demonstrates that typical evolutionary measures should not limit functional divergence to interaction sites between proteins. By combining phylogenetic analysis with molecular dynamics simulation, we were able to identify divergent residues in the TnI switch region and model their structural and functional effect. The interaction between teleost TnC1a/TnI1.5 is approximately half as energetically favourable as that between TnC1a/TnI1.1 or TnC1b/TnI1.1. The presence of divergent sites outside of this interface indicates that there is probably a further role for compensatory residue changes that occur between Tn subunits. Patterns in domain-specific divergent selection and in interaction energies suggest that substitutions in the TnI switch region are critical to modifying Tn function with paralog variation, providing evidence of structural sub-functionalization. This combination of phylogenetic inference with molecular dynamic modelling of structural interactions between paralogs profoundly strengthens our understanding of the molecular evolution of the Tn complex. The Tn complex can be an important model of the co-evolution of interacting proteins where functional divergence of paralog combinations guides variation in contractile function.
7.7. Acknowledgements

Thanks to Bruce Leighton for assistance with fish set-up and care, Haruyo Kashihara for assistance with fish sampling and Dr. Christopher Moyes of Queen’s University for careful reading of and input on the manuscript. This study was supported by grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada to GFT and to DPT and CIHR to GFT. GFT is a Tier I Canada Research Chair and DPT is an Alberta Innovates Health Solutions Scientist and Alberta Innovates Technology Futures Strategic Chair in (Bio)Molecular Simulations. Molecular dynamics simulations were carried out on the Westgrid Complex which is under the aegis of Compute Canada.
7.8. References


234


33. Brunet, N. M., Chase, P. B., Mihajlovic, G., and Schoffstall, B. (2014) Ca^{2+}-regulatory function of the inhibitory peptide region of cardiac troponin I is aided by the C-terminus of cardiac troponin T: Effects of familial hypertrophic cardiomyopathy mutations cTnI R145G and cTnT R278C, alone and in combination, on filament sliding. *Arch Biochem Biophys*


239


Figure S1: TnC maximum likelihood domain specific tree including only fish species.
The evolutionary history was inferred using the maximum likelihood method based on the K2 model for the MUSCLE alignment of TnC N-terminal domain (260 positions in final data set). The bootstrap consensus tree was inferred from 800 replicates with the percentage of replicate trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measure in the number of substitutions per site. The tree is rooted with lamprey TnC2. The analysis involved 89 nucleotide sequences of representative Actinopterygii TnC sequences compiled from NCBI and Ensembl. Evolutionary analyses were conducted in MEGA6.
Figure S2: TnI maximum likelihood domain specific tree including only fish species.
The evolutionary history was inferred using the maximum likelihood method based on the GTR+G model for the MUSCLE alignment of TnI functional domain encompassing IR/switch (108 nucleotide sites). The bootstrap consensus tree was inferred from 800 replicates with the percentage of replicate trees in which the associated taxa clustered together is shown next to the branches. The tree was drawn to scale, with branch lengths indicating the number of substitutions per site. The tree was rooted with lamprey TnI2. The analysis involved 128 nucleotide sequences of representative vertebrate TnI sequences compiled from NCBI and Ensembl. Evolutionary analyses were conducted in MEGA6.
Figure S3: RMSD plots as a function of time for replicated simulations of each paralog combination.
Each of the simulations has deviated from the starting coordinates between approximately 3 and 5 Å RMSD.
**Table S1:** Accession numbers for sequences used for a) TnC and b) TnI phylogenetic reconstruction.

<table>
<thead>
<tr>
<th>CODE</th>
<th>common name</th>
<th>species name</th>
<th>gene</th>
<th>accession</th>
<th>location</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZF</td>
<td>zebrafish</td>
<td><em>Danio rerio</em></td>
<td>TnC1a</td>
<td>GI:28822162</td>
<td>23:4946409-4951592:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnC1b</td>
<td>GI:50344823</td>
<td>23:20398636-20401146:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnC2</td>
<td>GI:41282127</td>
<td>23:20398636-20401146:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ENSDARG00000095002</td>
<td>23:20398636-20401146:1</td>
</tr>
<tr>
<td>RT</td>
<td>rainbow trout</td>
<td><em>Oncorhynchus mykiss</em></td>
<td>TnC1a</td>
<td>GI:30721848</td>
<td>23:20398636-20401146:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnC1b</td>
<td>contig59213</td>
<td>23:20398636-20401146:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnC2</td>
<td>197632073</td>
<td>23:20398636-20401146:1</td>
</tr>
<tr>
<td>AS</td>
<td>Atlantic salmon</td>
<td><em>Salmo salar</em></td>
<td>TnC1a</td>
<td>GI:209732707</td>
<td>23:20398636-20401146:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnC1b</td>
<td>GI:209737061</td>
<td>23:20398636-20401146:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnC2</td>
<td>197632073</td>
<td>23:20398636-20401146:1</td>
</tr>
<tr>
<td>BICHIR</td>
<td>gray bichir</td>
<td><em>Gasterosteus aculeatus</em></td>
<td>TnC1a</td>
<td>ENSGACG00000015148</td>
<td>scaffold_68:530265-533797:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnC1b</td>
<td>GI:112420155</td>
<td>groupXII:17267202-17271644:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnC2</td>
<td>ENSGACG00000004198</td>
<td>XII:3,692,240-3,694,594</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ENSGACG00000004196</td>
<td>XII:3,692,240-3,694,594</td>
</tr>
<tr>
<td>COD</td>
<td>Atlantic cod</td>
<td><em>Gadus morhua</em></td>
<td>TnC1a</td>
<td>GI:20386545</td>
<td>23:20398636-20401146:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnC1b</td>
<td>285333725</td>
<td>23:20398636-20401146:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnC2</td>
<td>ENSGMOG00000017790</td>
<td>23:20398636-20401146:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ENSGMOG00000017780</td>
<td>23:20398636-20401146:1</td>
</tr>
<tr>
<td>MD</td>
<td>Japanese medaka</td>
<td><em>Oryzias latipes</em></td>
<td>TnC1a</td>
<td>GI:187445749</td>
<td>7:18063144-18066928:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnC1b</td>
<td>187690733</td>
<td>scaffold541:131422-134224:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnC2</td>
<td>ENSORLG00000001797</td>
<td>7:4,282,424-4,285,314</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ENSORLG00000001761</td>
<td>7:4,282,424-4,285,314</td>
</tr>
<tr>
<td>FUGU</td>
<td>fugu</td>
<td><em>Takifugu rubripes</em></td>
<td>TnC1a</td>
<td>ENSTRUT00000013316</td>
<td>7:4,282,424-4,285,314</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnC2</td>
<td>ENSTRUG00000004150</td>
<td>7:4,282,424-4,285,314</td>
</tr>
<tr>
<td>TD</td>
<td>green pufferfish</td>
<td><em>Tetraodon fluviatilis</em></td>
<td>TnC1a</td>
<td>GI:28822168</td>
<td>23:20398636-20401146:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnC1b</td>
<td>AF89498.1</td>
<td>23:20398636-20401146:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnC2</td>
<td>197632073</td>
<td>23:20398636-20401146:1</td>
</tr>
<tr>
<td>EL</td>
<td>northern pike</td>
<td><em>Esox lucius</em></td>
<td>TnC1a</td>
<td>GI:225715250</td>
<td>23:20398636-20401146:1</td>
</tr>
<tr>
<td>SBF</td>
<td>sablefish</td>
<td><em>Anoplopoma fimbria</em></td>
<td>TnC1a</td>
<td>GI:229366608</td>
<td>23:20398636-20401146:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnC1b</td>
<td>229365730</td>
<td>23:20398636-20401146:1</td>
</tr>
<tr>
<td>TL</td>
<td>nile tilapia</td>
<td><em>Oreochromis niloticus</em></td>
<td>TnC1a</td>
<td>GI:348502953</td>
<td>GL831137.1:6592115-6595974:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnC1b</td>
<td>348533606</td>
<td>GL831137.1:6592115-6595974:1</td>
</tr>
<tr>
<td>Species</td>
<td>Genus</td>
<td>Species</td>
<td>Genus</td>
<td>GI:</td>
<td>GI:</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>-----------------</td>
<td>--------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>CATF blue catfish</td>
<td>Ictalurus</td>
<td>furcatus</td>
<td>Ictalurus</td>
<td>GL831362.1: 546,831-549,605</td>
<td>GL831362.1: 5538018-544902:1</td>
</tr>
<tr>
<td>channel catfish</td>
<td></td>
<td></td>
<td></td>
<td>TnC1a</td>
<td>GI: 204127059</td>
</tr>
<tr>
<td>mummichog</td>
<td>Fundulus</td>
<td>heteroclitus</td>
<td></td>
<td>TnC1a</td>
<td>GI: 224275321</td>
</tr>
<tr>
<td>TnC1b</td>
<td>GI: 55767539</td>
<td></td>
<td></td>
<td>TnC1b</td>
<td>GI: 48463171</td>
</tr>
<tr>
<td>MMC orange-spotted grouper</td>
<td>Epinephelus</td>
<td>coioides</td>
<td></td>
<td>TnC1a</td>
<td>GI: 295792342</td>
</tr>
<tr>
<td>GP rainbow smelt</td>
<td>Osmerus</td>
<td>mordax</td>
<td></td>
<td>TnC1a</td>
<td>contig2783</td>
</tr>
<tr>
<td>BT brook trout</td>
<td>Salvelinus</td>
<td>fontinalis</td>
<td></td>
<td>TnC1a</td>
<td>contig3129</td>
</tr>
<tr>
<td>WF lake whitefish</td>
<td>Coregonus</td>
<td>clupeaformis</td>
<td></td>
<td>TnC1a</td>
<td>Contig3006</td>
</tr>
<tr>
<td>BMB Burton's mouthbrooder</td>
<td>Haplochromis</td>
<td>burtoni</td>
<td></td>
<td>TnC1a</td>
<td>GI: 554857820</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC1b</td>
<td>GI: 554867010</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2.1</td>
<td>GI: 554855949</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GI: 554855947</td>
</tr>
<tr>
<td>AM Amazon molly</td>
<td>Poecilia</td>
<td>formosa</td>
<td></td>
<td>TnC1a</td>
<td>ENSPFOG00000009229</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC1b</td>
<td>ENSPFOG00000004162</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2.1</td>
<td>ENSPFOG0000019443</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ENSPFOG0000019446</td>
</tr>
<tr>
<td>MZ Marylandia zebra</td>
<td>zebra mbuna</td>
<td></td>
<td></td>
<td>TnC1a</td>
<td>GI: 498965643</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC1b</td>
<td>GI: 499040674</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2</td>
<td>GI: 499029626</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GI: 499029624</td>
</tr>
<tr>
<td>PN Red Mwanza</td>
<td>Pundamilia</td>
<td>nyerei</td>
<td></td>
<td>TnC1a</td>
<td>GI: 548483298</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC1b</td>
<td>GI: 548443920</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2</td>
<td>GI: 548437064</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GI: 548437058</td>
</tr>
<tr>
<td>NB Princess cichlid</td>
<td>Neolamprologus</td>
<td>brichardi</td>
<td></td>
<td>TnC1a</td>
<td>GI: 584003040</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC1b</td>
<td>GI: 584019322</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2</td>
<td>GI: 584018811</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GI: 584018809</td>
</tr>
<tr>
<td>ES Elephant shark</td>
<td>Callorhinchus</td>
<td>milii</td>
<td></td>
<td>TnC1a</td>
<td>GI: 632955498</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC1b</td>
<td>GI: 632980494</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2</td>
<td>GI: 632939530</td>
</tr>
<tr>
<td>DMF Bicolour damselfish</td>
<td>Stegastes</td>
<td>partitus</td>
<td></td>
<td>TnC1a</td>
<td>GI: 657582811</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC1b</td>
<td>GI: 657596608</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2</td>
<td>GI: 657566275</td>
</tr>
<tr>
<td>Species</td>
<td>Common Name</td>
<td>Genus</td>
<td>GI</td>
<td>Ensemblename</td>
<td>StartPosition</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------------------------</td>
<td>---------------------------</td>
<td>--------------</td>
<td>-------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>CF</td>
<td>cave fish</td>
<td><em>Astyanax mexicanus</em></td>
<td></td>
<td>TnC1a ENSAMG00000009117</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC1b ENSAMG00000013191</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2 ENSAMG000000012860</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2 ENSAMG00000012883</td>
<td></td>
</tr>
<tr>
<td>PF</td>
<td>platyfish</td>
<td><em>Xiphophorus maculatus</em></td>
<td></td>
<td>TnC1a ENSXMG00000017777</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC1b ENSXMG00000014969</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2 ENSXMG00000016572</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LY</td>
<td>Arctic lamprey</td>
<td><em>Lethenteron camtschaticum</em></td>
<td></td>
<td>TnC1a GI:2589016</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2 GI:2589013</td>
<td></td>
</tr>
<tr>
<td>ST</td>
<td>cloudy catshark</td>
<td><em>Scyliorhinus torazame</em></td>
<td></td>
<td>TnC1 St_CT0008272_b2</td>
<td></td>
</tr>
<tr>
<td>CM</td>
<td>chub mackerel</td>
<td><em>Scomber japonicus</em></td>
<td></td>
<td>TnC1b GI:336463064</td>
<td></td>
</tr>
<tr>
<td>JM</td>
<td>jack mackerel</td>
<td><em>Trachurus japonicas</em></td>
<td></td>
<td>TnC1b GI:406822720</td>
<td></td>
</tr>
<tr>
<td>SOLE</td>
<td>Senegalese sole</td>
<td><em>Solea senegalensis</em></td>
<td></td>
<td>TnC1b GI:213044187</td>
<td></td>
</tr>
<tr>
<td>FM</td>
<td>fathead minnow</td>
<td><em>Pimephales promelas</em></td>
<td></td>
<td>TnC1b GI:72732226</td>
<td></td>
</tr>
<tr>
<td>GUPPY</td>
<td>guppy</td>
<td><em>Gallus gallus</em></td>
<td></td>
<td>TnC1b GI:145875635</td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>chicken</td>
<td><em>Poecilia reticulata</em></td>
<td></td>
<td>TnC1 GI:136036</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2 GI:1311229</td>
<td></td>
</tr>
<tr>
<td>DD</td>
<td>opossum</td>
<td><em>Mondelphis domestica</em></td>
<td></td>
<td>TnC1 GI:126336323</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2 ENSMODG00000013668</td>
<td></td>
</tr>
<tr>
<td>LA</td>
<td>African elephant</td>
<td><em>Loxodonta africana</em></td>
<td></td>
<td>TnC1 GI:344276629</td>
<td></td>
</tr>
<tr>
<td>RAB</td>
<td>rabbit</td>
<td><em>Oryctolagus cuniculus</em></td>
<td></td>
<td>TnC1 GI:136040</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2 GI:136047</td>
<td></td>
</tr>
<tr>
<td>HS</td>
<td>human</td>
<td><em>Homo sapien</em></td>
<td></td>
<td>TnC1 GI:136038</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2 GI:4507617</td>
<td></td>
</tr>
<tr>
<td>MEG</td>
<td>megabat</td>
<td><em>Pteropus vampyrus</em></td>
<td></td>
<td>TnC1 ENSPVAG00000012568</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2 ENSPVAG0000000370</td>
<td></td>
</tr>
<tr>
<td>MM</td>
<td>mouse</td>
<td><em>Mus musculus</em></td>
<td></td>
<td>TnC1 GI:6678369</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2 GI:667837</td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>anole</td>
<td><em>Anolis carolinensis</em></td>
<td></td>
<td>TnC1 GI:327265791</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2 GI:46092546</td>
<td></td>
</tr>
<tr>
<td>XL</td>
<td>clawed frog</td>
<td><em>Xenopus laevis</em></td>
<td></td>
<td>TnC1 GI:1945537</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2 GI:7441462</td>
<td></td>
</tr>
<tr>
<td>SH</td>
<td>shrew</td>
<td><em>Sorex araneus</em></td>
<td></td>
<td>TnC1 ENSSARG0000000539</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TnC2 ENSTBEG0000002025</td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>Species</td>
<td>Common Name</td>
<td>Chromosome</td>
<td>Start</td>
<td>End</td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td>VV</td>
<td>Chlorocebus pygerythrus</td>
<td>vervet</td>
<td>11: 51,391,013-51,392,928</td>
<td>248</td>
<td>249</td>
</tr>
<tr>
<td>HY</td>
<td>Procavia capensis</td>
<td>rock hyrax</td>
<td>7: 32,510,878-32,512,309</td>
<td>249</td>
<td>250</td>
</tr>
<tr>
<td>EQ</td>
<td>Equus caballus</td>
<td>horse</td>
<td>1: 15,031,269-15,032,273</td>
<td>250</td>
<td>251</td>
</tr>
<tr>
<td>GP</td>
<td>Cavia porcellus</td>
<td>guinea pig</td>
<td>9: 15,437,550-15,438,647</td>
<td>251</td>
<td>252</td>
</tr>
<tr>
<td>TT</td>
<td>Tursiops truncatus</td>
<td>dolphin</td>
<td>3: 53,296,779-53,299,777</td>
<td>252</td>
<td>253</td>
</tr>
<tr>
<td>PIG</td>
<td>Sus scrofa</td>
<td>pig</td>
<td>2: 15,031,234-15,032,273</td>
<td>254</td>
<td>255</td>
</tr>
<tr>
<td>MG</td>
<td>Meleagris gallopavo</td>
<td>turkey</td>
<td>15: 32,994,757-32,995,847</td>
<td>255</td>
<td>256</td>
</tr>
<tr>
<td>ZF</td>
<td>Taeniopygia guttata</td>
<td>zebrafinch</td>
<td>1: 31,023,549-31,024,660</td>
<td>256</td>
<td>257</td>
</tr>
<tr>
<td>PT</td>
<td>Pan troglodytes</td>
<td>chimpanzee</td>
<td>11: 40,819,429-40,824,340</td>
<td>257</td>
<td>258</td>
</tr>
<tr>
<td>FLY</td>
<td>Ficedula albicollis</td>
<td>flycatcher</td>
<td>2: 15,031,234-15,032,273</td>
<td>258</td>
<td>259</td>
</tr>
<tr>
<td>CAT</td>
<td>Felis catus</td>
<td>cat</td>
<td>1: 31,023,549-31,024,660</td>
<td>259</td>
<td>260</td>
</tr>
<tr>
<td>PAN</td>
<td>Ailuroidea melanoleuca</td>
<td>panda</td>
<td>11: 40,819,429-40,824,340</td>
<td>260</td>
<td>261</td>
</tr>
<tr>
<td>TMNT</td>
<td>Pelodiscus sinensis</td>
<td>Chinese softshell turtle</td>
<td>1: 31,023,549-31,024,660</td>
<td>261</td>
<td>262</td>
</tr>
<tr>
<td>COEL</td>
<td>Latimeria chalumnae</td>
<td>coelecanth</td>
<td>11: 40,819,429-40,824,340</td>
<td>262</td>
<td>263</td>
</tr>
<tr>
<td>CODE</td>
<td>common name</td>
<td>species name</td>
<td>gene</td>
<td>accession</td>
<td>location</td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
<td>--------------</td>
<td>--------</td>
<td>-------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>RT</td>
<td>rainbow trout</td>
<td><em>Oncorhynchus mykiss</em></td>
<td>Tn1.1</td>
<td>NM_001185028</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn1.2</td>
<td>RTContig59584</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn1.3</td>
<td>BT057890</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn1.5</td>
<td>RTContig50917</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn1.6</td>
<td>BT057580</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn2.4</td>
<td>Contig40780</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn2.1</td>
<td>NM_001129990</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn2b</td>
<td>Contig34239</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn2.2</td>
<td>Contig54113</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn2.3</td>
<td>Contig48710</td>
<td></td>
</tr>
<tr>
<td>AS</td>
<td>atlantic salmon</td>
<td><em>Salmo salar</em></td>
<td>Tn1.1</td>
<td>SSContig79376</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn1.5</td>
<td>BT049066</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn1.6</td>
<td>BT057580</td>
<td></td>
</tr>
<tr>
<td>ZF</td>
<td>zebrafish</td>
<td><em>Dario rerio</em></td>
<td>Tn1.1</td>
<td>NM_001008613</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn1.2</td>
<td>ENSDARG000000028343</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn1.3</td>
<td>GI:50540325</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn1.4</td>
<td>NM_001045382.1.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn1.5</td>
<td>GI:323510638</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn1.6</td>
<td>NM_001002101</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn2a1</td>
<td>ENSDART00000033872</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn2a2</td>
<td>ENSDART00000149439</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn2a3</td>
<td>ENSDART00000012256</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn2a4</td>
<td>ENSDARG00000029069</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn2b1</td>
<td>ENSDARG00000035958</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn2b2</td>
<td>ENSDARG00000029995</td>
<td></td>
</tr>
<tr>
<td>CF</td>
<td>cavefish</td>
<td><em>Astyanax mexicanus</em></td>
<td>Tn1.1</td>
<td>ENSAMXG00000013570</td>
<td>KB871995.1:166509-173290:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn1.3</td>
<td>ENSAMXT00000005827</td>
<td>KB871833.1:2167929-2182452:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn1.4</td>
<td>ENSAMXT0000003991</td>
<td>KB882081.1:4023978-4027183:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn1.5</td>
<td>ENSAMXT000001441</td>
<td>KB882164.1:69,936-97,299</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn2a1</td>
<td>ENSAMXT0000017638</td>
<td>KB871579.1:4288239-4289716:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn2a2</td>
<td>ENSAMXT0000017633</td>
<td>KB871579.1:4280732-4287002:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn2a3</td>
<td>ENSAMXT0000017624</td>
<td>KB871579.1:4269007-4271694:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tn2a4</td>
<td>ENSAMXT0000017612</td>
<td>KB871579.1:42665250-4266985:1</td>
</tr>
<tr>
<td>Organism</td>
<td>Species</td>
<td>Tn1b1 Accession</td>
<td>Tn1a1 Accession</td>
<td>Tn1a2 Accession</td>
<td>Tn2a1 Accession</td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>cod</td>
<td>cod</td>
<td>Tn1.1 ENSGMOT00000011745</td>
<td>Tn1.2 ENSGMOT00000016662</td>
<td>Tn1.5 ENSGMOT0000001208</td>
<td>Tn2a1 ENSGMOT00000018580</td>
</tr>
<tr>
<td>fugu</td>
<td>fugu</td>
<td>Tn1.1 ENSTRU00000030805</td>
<td>Tn1.4 ENSTRU00000039948</td>
<td>Tn1.5 ENSTRU00000043333</td>
<td>Tn2a1 ENSTRU00000042707</td>
</tr>
<tr>
<td>MD</td>
<td>medaka</td>
<td>Tn1.1 ENSORLT00000008782</td>
<td>Tn1.2 ENSORL00000008327</td>
<td>Tn1.3 ENSORLT00000018387</td>
<td>Tn1.5 ENSORLT00000004577</td>
</tr>
<tr>
<td>PF</td>
<td>platfish</td>
<td>Tn1.1 ENSXMAT00000018101</td>
<td>Tn1.2 ENSXMAT0000008205</td>
<td>Tn1.3 ENSXMAT0000005457</td>
<td>Tn1.5 ENSXMAT0000017084</td>
</tr>
<tr>
<td>SB</td>
<td>stickleback</td>
<td>Tn1.1 ENSGACT00000006752</td>
<td>Tn1.2 ENSGACT00000012065</td>
<td>Tn1.3 ENSGACT00000026140</td>
<td>Tn1.4 ENSGACT00000017174</td>
</tr>
<tr>
<td>Species</td>
<td>Member</td>
<td>Assembly</td>
<td>Accession</td>
<td>Group</td>
<td>Coordinates</td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>----------</td>
<td>----------------------------</td>
<td>-----------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>TD tetraodon</td>
<td>TnI1.5</td>
<td>ENSGACT00000005583</td>
<td>groupXIX: 5,987,728-5,990,910</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI1.6</td>
<td>ENSGACT00000005601</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI2a1</td>
<td>ENSGACT00000011019</td>
<td>groupXIX: 9287888-9290073:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI2a2</td>
<td>ENSGACT00000011028</td>
<td>groupXIX: 9291925-9297476:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI2b1</td>
<td>ENSGACT00000022942</td>
<td>21883845:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI2.3</td>
<td>ENSGACT00000011051</td>
<td>groupXIX: 9320462-9322445:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI2.4</td>
<td>ENSGACT00000011084</td>
<td>9326473:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TL nile tilapia</td>
<td>TnI1.1</td>
<td>ENSTNIT00000018310</td>
<td>9:4624817-4625951:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI1.3</td>
<td>ENSTNIT00000014164</td>
<td>19:2872762-2874519:-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI1.5</td>
<td>ENSTNIG00000015998</td>
<td>13:10,162,092-10,163,564</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI2a1</td>
<td>ENSTNIT00000005522</td>
<td>13:13233445-13237809:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI2a2</td>
<td>ENSTNIT00000005521</td>
<td>13:13238586-13242701:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI2b1</td>
<td>ENSTNIT00000022215</td>
<td>5:12360348-12361683:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI2.1</td>
<td>ENSTNIT0000000712</td>
<td>13:13251994-13253139:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI2.3</td>
<td>ENSTNIT00000005519</td>
<td>13:13246817-13250014:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI2.4</td>
<td>ENSTNIG00000002812</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CATF channel catfish</td>
<td>TnI1.3</td>
<td>GI:318065094</td>
<td>GL31272.1:824836-828012:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI1.5</td>
<td>GI:308322430</td>
<td>GL31196.1:2458662-2463447:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GL31167.1:2013196-2018460:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GL31142.1:5136882-5140327:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GL31160.1:2,611,046-2,615,563</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI2a1</td>
<td>ENSONIT00000013559</td>
<td>GL311139.1:2397189-2398929:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI2a2</td>
<td>ENSONIT00000013561</td>
<td>GL31139.1:2409389-2411854:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI2.3</td>
<td>ENSONIT00000013564</td>
<td>GL31139.1:2422344-2425865:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI2.4</td>
<td>ENSONIT00000013570</td>
<td>GL31139.1:2430393-2431313:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI2.5</td>
<td>ENSONIT00000013571</td>
<td>GL31139.1:2435874-2438798:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI2.6</td>
<td>ENSONIT00000013572</td>
<td>GL31139.1:2441534-2444016:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZM zebra mbuna</td>
<td>TnI1.1</td>
<td>GI:498971303</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TnI1.2</td>
<td>GI:499016167</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Species</td>
<td>Organism</td>
<td>GI Numbers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------------------------</td>
<td>----------------------------</td>
<td>------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NB</td>
<td>Princess cichlid</td>
<td><em>Neolamprologus brichardi</em></td>
<td>TnI1.1 GI:583975044</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnI1.2 GI:548370811</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Red Mwanza</td>
<td><em>Pundamilia nyererei</em></td>
<td>TnI1.1 GI:554807661</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnI1.2 GI:554807659</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HB</td>
<td>Burton's cichlid</td>
<td><em>Haplochromis burtoni</em></td>
<td>TnI1.1 GI:597751541</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnI1.2 GI:597751543</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mexican tetra</td>
<td><em>Asyanax mexicanus</em></td>
<td>TnI1.1 GI:597751541</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnI1.2 GI:597751543</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnI1.3 GI:597745994</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TnI2 GI:597731604</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AM</td>
<td>Amazon molly</td>
<td><em>Poecilia formosa</em></td>
<td>KI519628.1:1165748-1169311:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>KI519627.1:176,740-177,745</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>KI519764.1:1020247-1025201:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>KI519627.1:178909-180870:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>KI519627.1:187357-190211:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>KI519834.1:472323-482716:-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>KI519709.1:1521116-1523899:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAR</td>
<td>Spotted gar</td>
<td><em>Lepisosteus oculatus</em></td>
<td>ENSLOC000000009801</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LG3:23976828-23983441:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LG8:45607031-45621369:-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>JHS940111.1:192,399-200,828</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LG8:45607031-45621369:-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LY</td>
<td>Lamprey</td>
<td><em>Petromyzon marinus</em></td>
<td>ENSPMAG00000000464</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XL</td>
<td>African clawed frog</td>
<td><em>Xenopus laevis</em></td>
<td>ENSXETG00000017087</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ENSXETG00000004621</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ENSXETG00000024676</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>Green lizard</td>
<td><em>Anolis carolinensis</em></td>
<td>ENSACAG00000003672</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4:131,460,969-131,469,155</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>Species</td>
<td>Genus</td>
<td>Species</td>
<td>Chromosome</td>
<td>Start</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>-------</td>
<td>---------</td>
<td>-------------</td>
<td>-------</td>
</tr>
<tr>
<td>GG</td>
<td>chicken</td>
<td>Gallus</td>
<td>gallus</td>
<td>26</td>
<td>774,174-777,490</td>
</tr>
<tr>
<td>HS</td>
<td>human</td>
<td>Homo</td>
<td>sapiens</td>
<td>TnI1</td>
<td>1:201,403,768-201,429,866</td>
</tr>
<tr>
<td>BT</td>
<td>cow</td>
<td>Bos</td>
<td>taurus</td>
<td>TnI1</td>
<td>16: 49,293,852-49,303,508</td>
</tr>
<tr>
<td>FD</td>
<td>cat</td>
<td>Felis</td>
<td>catus</td>
<td>TnI1</td>
<td>39,706,095-39,712,211</td>
</tr>
<tr>
<td>MM</td>
<td>mouse</td>
<td>Mus</td>
<td>musculus</td>
<td>TnI1</td>
<td>1: 135,779,434-135,810,989</td>
</tr>
<tr>
<td>RR</td>
<td>rat</td>
<td>Rattus</td>
<td>norvegus</td>
<td>TnI1</td>
<td>13: 57,672,836-57,685,255</td>
</tr>
<tr>
<td>VV</td>
<td>vervet</td>
<td>Chlorocebus</td>
<td>pygerythrus</td>
<td>TnI1</td>
<td>26: 27,997,709-28,002,646</td>
</tr>
<tr>
<td>ZBF</td>
<td>zebrafish</td>
<td>Aeolypgia</td>
<td>guttata</td>
<td>TnI1</td>
<td>26: 2,511,903-2,514,794</td>
</tr>
<tr>
<td>DD</td>
<td>oppossum</td>
<td>Monodelphis</td>
<td>domestica</td>
<td>TnI1</td>
<td>253</td>
</tr>
</tbody>
</table>

**GG** is the chicken (Gallus gallus) and **HS** is the human (Homo sapiens). The table lists the chromosome, start, and end positions for various genes and species.
### Table S2: Zebrafish primers used for quantitative PCR

<table>
<thead>
<tr>
<th>Accession number</th>
<th>qPCR probes</th>
<th>Amplicon size</th>
</tr>
</thead>
</table>
| ZF TnI1.1        | NM_001008613 | F: GCTGTTCGAGAACTACATGCT  
|                  |             | R: ACACCCGGACCTCCCTCTCAG  
|                  |             | 164            |
| ZF TnI1.2        |             | F: GTCAAGAGCTCATGAGTAAGATAGA  
|                  |             | R: CGACAGACACCTCCTTCCTCTCC  
|                  |             | 164            |
| ZF TnI1.3        | GI:50540325  | F: CAACCTTTATCCAAAAATGACACAGTA  
|                  |             | R: AGCAGATAACAGCACTCCTCC  
|                  |             | 159            |
| ZF TnI1.4        | NM_001045382 | F: CACATGTCGAGAATCATCTCTCT  
|                  |             | R: TCTCCACGGCTGTTGCTCTT  
|                  |             | 156            |
| ZF TnI1.5        | GI:323510638 | F: AGATTTCACACCAAGATTGA  
|                  |             | R: ATGCGCTCTGAGCTCTCTCT  
|                  |             | 168            |
| ZF TnI1.6        | NM_001002101 | F: CAAGACCTTCAGCTCTGCTGC  
|                  |             | R: TCTGCTACGGGTGCGCTCT  
|                  |             | 168            |
| ZF TnI2a1        | ENSDART00000033872 | F: GGAACGTGCAGAGATCTCTCA  
|                  |             | R: GACATGCGACTTTTCCTCAG  
|                  |             | 168            |
| ZF TnI2a2        | ENSDART00000149439 | F: AAGACATTGTCCCTGACTTTCA  
|                  |             | R: TCCTGACCTTATTGACAACA  
|                  |             | 148            |
| ZF TnI2a3        | ENSDART00000012256 | F: CAGGAGCTGTGCAAGAAAATCT  
|                  |             | R: CTTCTCAGGGAGCAGCTTC  
|                  |             | 160            |
| ZF TnI2a4        | ENSDARG00000029069  | F: TCTGCTGAGTCTGCTGCA  
|                  |             | R: CTCTTCAAGGGAGGCTTCTC  
|                  |             | 170            |
| ZF TnI2b1        | ENSDARG00000035958  | F: ACTATGTAGGAAGATGCACCA  
|                  |             | R: AGAGACATACGACACTTTCC  
|                  |             | 169            |
| ZF TnI2b2        | ENSDARG00000029995  | F: CTCAGTGACAGAAATACAGGA  
|                  |             | R: TCAGAGCAGGCTTCTTGAAT  
|                  |             | 170            |
| ZF b-actin       | AF057040      | F: TCTGTCGTTCTACTGTTGATTTGTG  
|                  |             | R: ATCTTCACTGAGTAGTCTGCTGG  
|                  |             | 170            |
| ZF ef1a proximal | xxxxxx        | F: GGAATTCGAGAGCACA  
|                  |             | R: GGCTCTGCTGGTCTCTGGAGA  
|                  |             | 168            |
| ZF ef1a distal   | xxxxxx        | F: AGATAACCGCAAGGAAGGT  
|                  |             | R: TCTAAGGGCACTGCTCCAC  
|                  |             | 175            |
### Table S3: Results of Branch-Specific Models testing for variation in ω-ratios among branches of full length TnC and N-terminal TnC in fish. Significance for the LRT was determined with p<0.0001.

<table>
<thead>
<tr>
<th>Model</th>
<th>LnL</th>
<th>kappa</th>
<th>ω</th>
<th>LRT test result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free-ratio</td>
<td>-12196.93</td>
<td>1.35</td>
<td>ω = 0.032</td>
<td></td>
</tr>
<tr>
<td>TnC fish specific full length 89 sequences 483 sites</td>
<td>-14605.32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-ratio (null)</td>
<td>-12196.93</td>
<td>1.35</td>
<td>ω = 0.032</td>
<td>ω(a) = 0.001</td>
</tr>
<tr>
<td>Two-ratios, branch 1 Prediction - positive selection on TnC1a</td>
<td>-12196.93</td>
<td></td>
<td></td>
<td>p&lt;0.0004</td>
</tr>
<tr>
<td>Two-ratios, branch 1 (null)</td>
<td>-12196.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two-ratios, branch 2 Prediction - positive selection on TnC1b</td>
<td>-12196.72</td>
<td>1.35</td>
<td>ω = 0.032</td>
<td>ω(b) = 0.018</td>
</tr>
<tr>
<td>Two-ratios, branch 2 (null)</td>
<td>-12196.72</td>
<td></td>
<td></td>
<td>p = 0.03</td>
</tr>
</tbody>
</table>

### Table S4: Results of Branch-Specific Models testing for variation in ω-ratios among branches of full length TnI and switch region specific TnI in fish. Significance for the LRT was assessed with p<0.0001.

<table>
<thead>
<tr>
<th>Model</th>
<th>LnL</th>
<th>kappa</th>
<th>Ω</th>
<th>LRT test result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free-ratio</td>
<td>-22612.80</td>
<td>1.51</td>
<td>ω = 0.10</td>
<td></td>
</tr>
<tr>
<td>TnI fish specific full length 128 sequences 363 sites</td>
<td>-24985.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-ratio (null)</td>
<td>-22612.22</td>
<td>1.51</td>
<td>ω(1.1) = 0.051</td>
<td></td>
</tr>
<tr>
<td>Two-ratios, branch 1 Prediction - positive selection on TnnI1.1</td>
<td>-22610.49</td>
<td>1.51</td>
<td>ω(1.5) = 0.94</td>
<td>1.57</td>
</tr>
<tr>
<td>Two-ratios, branch 1 (null)</td>
<td>-22610.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two-ratios, branch 2 Prediction - positive selection on TnnI1.5</td>
<td>-22610.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two-ratios, branch 2 (null)</td>
<td>-22610.49</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Free-ratio | -4247.44 | 1.99  | ω = 0.049 |
| TnI fish specific IR/switch 128 sequences 108 sites | -5062.19 |       |       |                 |
| One-ratio (null) | -4246.63 | 1.99  | ω(1.1) = 0.15  | 4.45 | p<0.06 |
| Two-ratios, branch 1 Prediction - positive selection on TnnI1.1 | -4248.04 |       |       |                 |
| Two-ratios, branch 1 (null) | -4245.92 | 1.99  | ω(1.5) = 0.84  | 0.015 | p = 0.16 |
| Two-ratios, branch 2 Prediction - positive selection on TnnI1.5 | -4246.04 |       |       |                 |
| Two-ratios, branch 2 (null) | -4246.04 |       |       |                 |
### Table S5. MM/PBSA binding energies for TnC/TnI paralog combinations

<table>
<thead>
<tr>
<th></th>
<th>TnC1a/TnI1.1</th>
<th>SEM</th>
<th>TnC1a/TnI1.5</th>
<th>SEM</th>
<th>TnC1b/TnI1.1</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM-VDW</td>
<td>-2 x 10^3</td>
<td>6 x 10^1</td>
<td>-2 x 10^3</td>
<td>7 x 10^1</td>
<td>-2 x 10^3</td>
<td>7 x 10^1</td>
</tr>
<tr>
<td>MM-ELECTROSTATICS</td>
<td>-2 x 10^3</td>
<td>4 x 10^1</td>
<td>-1 x 10^3</td>
<td>3 x 10^1</td>
<td>-2 x 10^3</td>
<td>4 x 10^2</td>
</tr>
<tr>
<td>APBS-POLAR</td>
<td>7 x 10^3</td>
<td>2 x 10^1</td>
<td>7 x 10^3</td>
<td>3 x 10^1</td>
<td>6 x 10^3</td>
<td>2 x 10^2</td>
</tr>
<tr>
<td>APBS-SASA</td>
<td>-2 x 10^3</td>
<td>9 x 10^0</td>
<td>-3 x 10^2</td>
<td>9 x 10^0</td>
<td>-3 x 10^1</td>
<td>9 x 10^0</td>
</tr>
<tr>
<td>TOTAL</td>
<td>-1 x 10^7</td>
<td>3 x 10^2</td>
<td>-6 x 10^7</td>
<td>2 x 10^1</td>
<td>-1 x 10^3</td>
<td>3 x 10^2</td>
</tr>
</tbody>
</table>

### Table S6. Quality indicators for simulated structures

<table>
<thead>
<tr>
<th></th>
<th>TnC1a/TnI1.1</th>
<th>TnC1a/TnI1.5</th>
<th>TnC1b/TnI1.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAMPAGE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored</td>
<td>96%</td>
<td>94%</td>
<td>96%</td>
</tr>
<tr>
<td>Allowed</td>
<td>4%</td>
<td>6%</td>
<td>4%</td>
</tr>
<tr>
<td>Disallowed</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Molprobity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bad Backbone Bonds</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Bad Backbone Angles</td>
<td>4.11%</td>
<td>5.33%</td>
<td>5.45%</td>
</tr>
<tr>
<td>Molprobity Score</td>
<td>1.65</td>
<td>1.97</td>
<td>2.31</td>
</tr>
<tr>
<td>QMEAN score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.744</td>
<td>0.649</td>
<td>0.688</td>
</tr>
<tr>
<td>Whatcheck structure Z-score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st generation packing quality</td>
<td>-0.84</td>
<td>-1.823</td>
<td>-1.601</td>
</tr>
<tr>
<td>2nd generation packing quality</td>
<td>0.286</td>
<td>-1.037</td>
<td>0.424</td>
</tr>
<tr>
<td>Ramachandran plot appearance</td>
<td>-1.415</td>
<td>-2.089</td>
<td>-2.172</td>
</tr>
<tr>
<td>$\chi_1/\chi_2$ rotamer normality</td>
<td>-3.272</td>
<td>-3.661</td>
<td>-3.229</td>
</tr>
<tr>
<td>Backbone conformation</td>
<td>1.024</td>
<td>0.627</td>
<td>1.232</td>
</tr>
<tr>
<td>Inside/outside distribution</td>
<td>0.959</td>
<td>1</td>
<td>1.077</td>
</tr>
<tr>
<td>Whatcheck RMS Z-score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths</td>
<td>0.352</td>
<td>0.344</td>
<td>0.343</td>
</tr>
<tr>
<td>Bond angles</td>
<td>1.985</td>
<td>1.997</td>
<td>2.004</td>
</tr>
<tr>
<td>Omega angle restraints</td>
<td>1.638</td>
<td>1.744</td>
<td>1.699</td>
</tr>
<tr>
<td>Side chain planarity</td>
<td>3.792</td>
<td>3.877</td>
<td>3.339</td>
</tr>
<tr>
<td>Improper dihedral distribution</td>
<td>2.192</td>
<td>2.256</td>
<td>2.072</td>
</tr>
</tbody>
</table>
Chapter 8.
General Discussion

Troponin (Tn) genes are highly conserved, which makes them strong candidates to investigate for duplicate retention after WGD (13, 52, 113). Teleost fish represent the most diverse vertebrate species, with the teleost-specific genome duplication (TS-WGD) an important event in generating teleost complexity. While the TS-WGD was not the direct factor generating teleost diversity, it is a critical factor in allowing for adaptation to changes in environmental conditions (48). In this dissertation, the sub-functionalization of Tn paralogs is used to provide insight into how teleosts achieve physiological versatility in contractile properties. Combining knowledge of both regulatory and structural coevolution patterns between paralogs is a novel approach to the use of the Tn complex as an evolutionary model.

8.1. Aim 1: The nature of the variation in cardiac function in teleosts

8.1.1. Variation in contractility is linked to mixed expression patterns of paralogs

The ability to vary cardiac contractility in response to environmental challenges is key for the survival of ectotherms such as the zebrafish. In our echocardiographic studies, cardiac output was modulated primarily by heart rate in response to acute temperature changes. Increased reliance on atrial contraction for ventricular filling is evidence of further remodeling due to cold acclimation. Alterations in the paralog composition of contractile protein complexes may be key to the maintenance of cardiac functional parameters in zebrafish when they are faced with environmental challenges.

In mammals, many sarcomeric proteins have muscle-specific paralogs that are expressed such that a single paralog is uniquely found in each muscle type. Both TnI and
TnT seem to follow this pattern in adult mammals (65). However, it is increasingly clear with exploration of basal vertebrates such as teleosts that Tn paralog expression patterns may be more complicated. Within cardiac muscle of both zebrafish and rainbow trout, overlapping patterns of expression are found for paralogs of TnC and TnI (2, 42, 117). In mammalian cardiac cells co-expression of multiple paralogs or splice variants in the Tn complex can result in modified protein interactions and variation in Ca\(^{2+}\) sensitivity leading to non-uniform contraction (128). Co-expression for Tn isoforms is often indicative of a diseased state of the mammalian heart (6, 128). For example, the co-presence of multiple TnT splice variants in adult mammalian hearts reduces cardiac efficiency by desynchronizing the Ca\(^{2+}\) activation of the thin filaments (76). Uniform function of contractile units is important for rhythmic function of the vertebrate heart (50) whereas in skeletal muscle a broader range of twitch strengths introduced by multiple TnT variants sum into greater force of tetanic contractions (50). While mixed isoform usage of both TnT and TnI is found during fetal cardiac development, co-expression results in less contractile force (33, 131). In ectothermic fish, the presence of multiple TnC paralogs with minimal differences in Ca\(^{2+}\) sensitivity may allow flexibility in response to variable environmental conditions while still maintaining uniform cardiac contraction. The expression patterns of both TnC and TnI in model teleosts that are examined in this dissertation change in response to acclimation temperature. However, fish must also be prepared for acute temperature fluctuations. The work presented in this dissertation suggests that mixed Tn complexes allow for increased tolerance to acute temperature change through graded contractile responses without transcriptional remodeling. A transcriptional response is found with acclimation but greater tolerance to short-term fluctuations can be created through the presence of mixed paralog complexes of contractile proteins with subtle differences in Ca\(^{2+}\) sensitivity.

While this dissertation focuses on the Tn complex in relation to overall contractility changes in the heart, it is difficult to attribute the overall functional changes that are associated with acclimation to a single protein. With the increasing use of zebrafish as a model of vertebrate cardiac physiology (105), the variation in paralog use must be considered due to the increased number of teleost-specific paralogous genes that code for cardiac proteins. Subunit composition of ion channels creates variation in ionic currents (125) and variation in contractile proteins can create variation in the mechanics of cardiac
contraction (78). Fish exhibit multiple orthologs of sarcomeric proteins relative to mammals, and the temperature tolerance of the myocardium is generated through paralog switches for these key contractile proteins (104). Expression of genes that code for proteins involved in protein synthesis and ATP-generating processes are typically adjusted with thermal acclimation of fish, as well as genes involved with muscle hypertrophy such as myosin light chain 2 (72, 81, 126). Multiple homologues of myosin heavy chain and actin are expressed in chamber-specific manner in adult zebrafish heart (117). Other examples in the zebrafish heart include remodeling of collagen content of the ventricle in order to maintain proper compliance for contraction upon cold acclimation (77). Decreased compliance of the zebrafish heart is supported by the observation of increased isovolumic relaxation time (IVRT) seen with cold acclimation in our echocardiography study, demonstrating longer time necessary for the ventricle to relax. The mixed expression of multiple contractile-protein paralogs within a tissue may be an important part of teleost robustness to seasonal environmental changes that does not sacrifice plasticity. The Tn complex plays a critical role in the regulation of cardiac contraction as a molecular switch. The work presented here suggests that variation in the paralog composition of Tn complexes is an example of how teleosts can maintain cardiac function in response to changing environmental conditions.

In fish, modulation of the contractile properties that underlie the active contraction of the atrium helps to maintain cardiac output at different temperatures. The WA zebrafish in our echocardiography study showed broader tolerance in response of cardiac output to acute temperature change (Q10 of 2.15 in WA vs 1.89 in CA). An ectothermic fish will have a greater difficulty with the maintenance of CO with an increase in temperature relative to a decrease in temperature (32). This dissertation suggests that the reason that WA fish are better able to cope at both temperatures is due to mixed paralog composition of contractile proteins. Tn complexes composed of a mixture of paralogs, as we have found in warm acclimated fish, may confer greater phenotypic breadth in response to acute temperature change. At warm temperatures, the ventricle would be expected to have an increased relaxation rate based on the lower interaction strength that this work describes for the Tn complex that contains the TnI1.5 paralog and the preferential expression of TnI1.5 in the WA ventricle. A greater change in Ca2+ sensitivity would be conferred due to substitution of TnI paralogs than TnC paralogs. Increased tolerance of the ventricle to
environmental fluctuations is more important at warmer temperatures in zebrafish based on TnI paralog expression.

The echo data, presented in Chapter 4, demonstrates decreased early peak (E) velocity and decreased early peak velocity / atrial peak velocity (E/A) ratio in cold-acclimated fish. A decreased E/A ratio suggests increased reliance on atrial contraction for ventricular filling with cold acclimation, and decreased reliance on the atrium with warm acclimation in zebrafish. In teleosts, active contraction of the atrium is the major driving force of ventricular filling rather than central venous pressure (20). Anatomical differences in wall thickness of the atrium and ventricle may drive differences in force generation such that individual atrial cardiomyocytes produce greater maximal force (116). The atrium has a larger SR store of Ca\(^{2+}\) than the ventricle (1, 60). Lower RYR sensitivity to cytosolic Ca\(^{2+}\) enables higher SR Ca\(^{2+}\) loads by reducing SR Ca\(^{2+}\) release (60). Atrial muscle of fish also demonstrates higher sarcoplasmic reticulum Ca\(^{2+}\)-ATPase activity (SERCA2a), which contributes to increased SR Ca\(^{2+}\) load (1). The lower responsiveness of RyR clusters to Ca\(^{2+}\) concentration is due to lower density and greater spatial separation of individual ryanodine clusters in atrial cells (115). The PLN/SERCA ratio also shows differential expression in both the zebrafish and rainbow trout atrium, which is favorable for increased activity of the SR Ca\(^{2+}\) pump which results in the faster Ca\(^{2+}\) reuptake necessary for faster relaxation (51, 82). Lower responsiveness to cytosolic Ca\(^{2+}\) concentration delays Ca\(^{2+}\) induced Ca\(^{2+}\) release (CICR), which helps to prevent arrhythmia in atrial cells. The loss of function of NCX in mutant zebrafish (tre mutant) produces a greater incidence of cardiac fibrillation in the atrium and prevents contraction in the ventricle (27, 86), The atrium may be more susceptible to spontaneous Ca\(^{2+}\) release from the SR, but may also be better than the ventricle at utilizing other forms of Ca\(^{2+}\) extrusion for adaptive compensation (27). The increased Ca\(^{2+}\) sensitivity and greater Ca\(^{2+}\) buffering capacity in the atrium conferred by the atrium-specific TnC paralog substitution may function to promote a greater delay in ventricular contraction following atrial contraction, further working to decrease atrial arrhythmogenic activity. These atrial-specific adaptations contribute to stable SR function under widely variable temperatures (60), especially considering that the higher heart rates precipitated by warmer temperatures increase the risk of arrhythmogenesis (60, 86, 94, 127).
Gene expression of TnC and TnI fluctuate with temperature. Importantly, variation in TnC and TnI expression is observed between diverse teleost species. It is unclear whether differential paralog usage is a consequence of different environments inhabited by zebrafish and trout or a consequence of diverse phylogenetic histories. Teleosts at similar temperatures do not have similar expression profiles for contractile proteins, as trout hearts display a much greater TnC1a/TnC1b ratio in both chambers at 15ºC than zebrafish at 18ºC. Zebrafish hearts maintain absolute TnC1a transcript levels and modulate TnC1b levels at both temperatures, while in trout both TnC1a and TnC1b expression changes with temperature condition. The function of many proteins, such as Ca²⁺ binding parvalbumin, has been shown to vary more as a function of phylogenetic distance across species than by native temperature in teleosts (31). Any changes to sarcomeric proteins that encourage stabilization of atrial function are likely only be necessary in fish that live at warm temperatures, such as zebrafish (120). The residues in zebrafish TnC1b that create divergence of Ca²⁺ binding properties in zebrafish are not conserved across teleosts. Further characterization of Tn expression patterns is necessary to determine the extent to which paralog usage is influenced by environment. This characterization would include comparison of species with closer phylogenetic history, such that warm water salmonids are compared with trout, and cold water cyprinids are compared with zebrafish.

8.2. Aim 2: The context-specific sub-functionalization of the cardiac-specific TnC genes.

8.2.1. Gene duplication in driving sub-functionalization

Whole genome duplication events such as the teleost-specific 3R introduce the ability to duplicate entire pathways, enabling pathway-level innovation and retention of interacting partners. This process as a whole is advantageous for introducing evolutionary novelty. Genetic redundancy can also lend robustness to fish, allowing teleosts to maintain a stable phenotype under environmental challenges. However, products retained from 3R may be mixed with products of tandem gene duplication, such as tandem duplicate TnC
paralogs that interact with 3R derived TnI paralogs. This means that new functions at the individual protein level are coupled to the cooperation between interacting partners. It is important to understand that partners not only have biochemical interaction but also genetic interactions. This dissertation discusses selection as it happens on an individual protein level as well as the equally important selection upon correlated properties of interacting proteins in the process of sub-functionalization.

8.2.2. Inter-relation of regulatory sub-functionalization and structural sub-functionalization

In these studies we found evidence of both regulatory sub-functionalization of TnC1 paralogs in the form of transcriptional expression patterns, as well as structural sub-functionalization in the form of changes in tertiary structure and Ca$^{2+}$ binding abilities of each TnC paralog. Functional divergence typically refers to changes in coding sequence, and is often thought of as changes on the biochemical level; however changes are observed at the regulatory level, coding level, and post-translational level. Molecular mechanisms of functional change at the regulatory level include changes in transcriptional expression of paralogous proteins (38, 41, 62). Changes at the coding level appear as changes in specificity or activity between paralogs (19), while changes at the posttranslational level manifest as modifications in regulation and localization (106). When referring to functional divergence the focus should not be limited to the change in proteins coded for by divergent genes, but also include the evolution of the way these genes are expressed. A strong positive correlation has been found between expression level and duplicate gene retention, which suggests a dosage response of retained duplicates or improved fitness with increased amounts of protein products (13, 52). The fact that retained duplicates are expressed differentially underscores the importance of determining the functional regulation of products created by duplication events. Transcriptional changes have been established as major contributors to functional change in duplicates (26, 56). Whether functional divergence occurs predominantly by changes in gene regulation or by changes within the amino acid coding sequence of the proteins is not yet clear (92).
Changes in transcriptional expression arise in response to changes in the regulatory sequences of genes. DNA binding site motifs regulate function by providing transcription factors the opportunity to bind to genomic elements and affect the expression of nearby genes. Evolutionary changes to functional DNA elements are hypothesized to be major contributors to phenotypic diversity within and between species (69). Cis-regulatory elements are small and interspersed with non-relevant sequences, often far away from the regulated gene and less conserved in sequence, which makes regulatory changes found in cis-elements difficult to detect. Their position can be changed or inverted without affecting their function, which makes these areas much harder to identify (129). Motifs may not be conserved across phylogeny as transcription factors are often products of sub-functionalization themselves (83). Most binding site prediction software is based upon mammalian TF binding, which adds increased difficulty in the study of promoters in lower vertebrates such as the teleosts focused on in this study.

The gene expression of duplicate genes can be divided between tissues or developmentally by gain or loss of cis-regulatory elements. In teleosts specifically, expression patterns between developmental stages often diverge between paralogs that result from TS-WGD and have undergone neo- and sub-functionalization (79). The consequences of cis-regulatory sequence changes are often identified through changes in expression analyses, such as quantitative PCR, as used in this thesis for TnC and TnI paralogs. Differential expression suggests differences in regulation of expression. In the mammalian heart the critical transcription factors GATA4 and MEF2c bind to the promoter regions of several genes including both TnC1 and TnI3 (70, 99, 103). The co-occurrence of multiple cardiac transcription factors is required for cardiac transcriptional activity (61), but the extent to which they would influence expression and timing of expression of Tn paralogs in fish is unknown. Without further information on the regulation of transcriptional difference between paralogs we cannot comment on whether co-expression of TnC/TnI paralogs is due to cis-regulation.

The importance of gene expression and coding sequence/function evolution is underlined by the progressive nature of evolution. The DDC model suggests that sub-functionalization leads to neo-functionalization (79). The isolation of genes by first changing thier non-coding sequence releases constraints on their products through the
development of independent regulatory patterns that allows new functions to evolve. Through sub-functionalization of coding sequence, duplicates satisfy the functional requirements of the ancestral locus. Yet, gene duplication produces a dosage imbalance in the encoded proteins, which is resolved by diversification of expression patterns between paralogs (34). The “duplication degeneration innovation” model (DDI) suggests that sub- and neo-functionalization act on regulatory elements to introduce novelty (75). After duplication and during degeneration, non-coding regulatory elements are either conserved or become non-functional between the duplicates in a way that increases the specialization of expression for each duplicate gene (79). Any novel spatial and/or temporal expression patterns can be achieved even by only subtle mutations in degenerate regulatory regions, which makes neo-functionalization more likely to occur between duplicates. Pleiotropic, multifunctional genes with complex regulatory sequences typically have an increased chance of sub-functionalization with duplication (79). Genes that code for proteins with multiple roles or express in multiple tissues, like TnC1 or TnI1, have a greater chance of dividing those roles among duplicates. While this model predominantly focuses on the evolution of cis-regulation, it can further be expanded to changes observed in coding sequences.

The distribution of an ancestral function onto two paralogous genes can also be achieved by reciprocal inactivation of functional domains in coding sequences. A new function can be gained through the addition of another regulatory element, which results in a new expression domain, or through imposing a new function via changes within the coding sequence. While changes in coding sequences are normally thought to result from point mutations, divergence can also result from insertions/deletions (indels), exon gain/loss, exonization/pseudoxonization, exon shuffling and divergence of alternative splicing (10, 64, 74). These multiple mechanisms of creating divergence reiterate the importance of understanding functional divergence at multiple levels rather than simply through coding sequence. The coupling of changes in expression and function have been shown where variation of TF binding sites is correlated with coding sequence substitution between duplicates (5). Both non-coding and coding sequences of certain paralogs in teleosts have been shown to undergo divergent evolution. In particular, both scenarios of sub-functionalization have been shown for proopiemiaocortins in pufferfish, prohormones that are primarily expressed in the pituitary gland (24), and follistatins in
salmonids, TGF-β binding proteins that are involved in muscle development (100). The retention of teleost duplicates for cardiac proteins such as Tn subunits may also have involved a combination of transcriptional regulation and structural changes given the predicted phylogenetic time since the duplication event.

Whether functional divergence occurs predominantly by changes in gene regulation or by changes within the amino acid coding sequence of the proteins is not yet clear (92). Looking at Tn functional divergence based solely on coding sequence or structural sub-functionalization is insufficient, as these proteins accumulate fewer mutations that do not have detrimental effects on function (23, 113). Hence we approach functional divergence looking at the inter-relatedness of coding and non-coding sequence evolution in the sub-functionalization process (93, 98, 129). The mixed expression of transcripts for both TnC and TnI in teleosts rather than the simple “one paralog per tissue” pattern suggests an overlapping function, that has been optimized through differential expression of each paralog in a temperature- and chamber-specific way. The functional difference in the interaction of TnI and TnC is created through variation in the paralog composition of the Tn complex, and reveals the importance of both types of sub-functionalization in shaping Tn in teleosts.

8.2.3. Structural sub-functionalization of TnC paralogs

The retention of genes due to variation in regulation of expression alone is difficult to contextualize unless accompanied by coding sequence variation that changes function of the protein in some way. However, defining structural sub-functionalization simply as changes in coding sequence is misleading – especially in proteins as well characterized as TnC. Many studies based on genes products lack structural information that hinders phylogenetic analysis, but protein structure alone without phylogenetic information may not allow for sufficient detection of evolutionary divergence. Neo-functionalization and sub-functionalization are important processes that facilitate the retention of duplicate genes, and consideration of structural constraints placed upon genes that encode proteins aids our understanding of these processes (112). Specific positions in protein-encoding genes are under varying selection pressures, making it necessary to understand the complex
dynamics in a folded protein (111). The relative importance of specific residues in an interaction with a ligand or another protein are reflected in the degree of conservation (123). The more important the residue, the sooner it becomes “fixed” in different evolutionary branches, and the more divergent the branches become. A change in function may result even from a conservative amino acid change in an important residue. However, many mutations are fixed without an immediately apparent effect, as described by Kimura’s neutral theory (80). Many mutations can appear neutral in a given environment but can exhibit differential function in response to environmental change (102). Functionally important sites in protein families can be predicted based on the level of conservation by considering sequence and structure information. Identification of residues that define variation between paralogs of the troponin family is a measure of functional divergence that must be substantiated using structural and functional studies.

Typically, based on the Ca\(^{2+}\) handling properties of TnC, we would expect functional changes to manifest as increases or decreases in Ca\(^{2+}\) binding affinity, which in turn determine the Ca\(^{2+}\) sensitivity of the entire sarcomere. However, the Tn protein complex has a complex functional landscape that is not limited to Ca\(^{2+}\) binding. It is important to distinguish that changes in direct Ca\(^{2+}\) affinity due to variation in the EF hand motif are not the only way to change the function of Tn. Ca\(^{2+}\) sensitivity refers to the propensity of a muscle contraction to be triggered by a particular concentration of Ca\(^{2+}\). This is influenced by a myriad of modifiers between site II of cTnC and the myosin/actin interaction, including the interfaces between Tn complex members, tropomyosin, actin, myosin binding protein C, and myosin. In TnC1, Ca\(^{2+}\) affinity refers directly to the ability of TnC to bind Ca\(^{2+}\). The association of Ca\(^{2+}\) to this motif is dependent on the coordinating residues and also the overall conformation of the protein. There is evidence that this interaction is modified allosterically through changes in the interactions with other proteins. However, many sequence substitutions in sarcomeric proteins that have a functional effect are quite distant from this area and likely play a more prominent role in modulating myofilament sensitivity rather than TnC Ca\(^{2+}\) affinity directly.

The divergence between TnC2 and TnC1 in mammals provides an example of sequence modification of the EF-hand motif that directly results in variation in Ca\(^{2+}\) affinity. Site I of all TnC1 orthologs displays lower sequence similarity than TnC2, with sequence
substitutions that resulted in the inability of site I to bind Ca\textsuperscript{2+}(43). Site II is conserved in all TnC1 homologs examined to date suggestive of no direct variation in Ca\textsuperscript{2+} coordination (47). However, variation exists in residues surrounding Site I within vertebrate TnC1 orthologs, implicated in salmonid TnC1a, were implicated in creating functional divergence by increasing Ca\textsuperscript{2+} sensitivity of force production in a single-cell skinned mammalian cardiomyocyte preparation (44). However, reducing the variation between mammalian and teleost TnC1 to four residues (Asn2, Ile28, Gln29, and Asp30 = NIQD) was difficult to explain in terms of structural divergence due to minimal structural differences detected between the isolated N-terminal region of human cTnC and that with the NIQD substitutions introduced (133). None of the variation between TnC1a and TnC1b occurs in site II, and no direct effect on Ca\textsuperscript{2+} coordination would be expected. Much of the variation that distinguishes the TnC1 paralogs is located in the N-helix of TnC. This region is thought to be important for the stabilization of the relative orientation of TnC when in complex with TnI and TnT (9). While many of the variation in residues between paralogs are conservative in nature and would have minimal impact, the TnC1b paralog is missing one of the four residues of the fish-specific NIQD. This suggests that there should be a difference in Ca\textsuperscript{2+} sensitivity between TnC paralogs, but on a much lower scale than that has been found between mammals and fish TnC1a, especially when looking at TnC in isolation.

At lower temperatures Ca\textsuperscript{2+} sensitivity of the contractile element is decreased (17, 59). Higher relative TnC Ca\textsuperscript{2+} affinity in fish allows for proper cardiac function in colder environments (44, 46). MD simulation found little change in Ca\textsuperscript{2+} affinity between the two ZF TnC1 paralogs at both 18\textdegree C and 28\textdegree C. When measured by ITC, TnC1b had higher Ca\textsuperscript{2+} affinity. Ligand binding properties are conserved in many orthologous enzymes, when compared at physiological temperatures (37). Thermal adaptation of lactate dehydrogenase (LDH) enzymes demonstrate how flexibility changes affect mobility of adjacent active-site structures to ensure temperature-compensation for rates of catalysis (37). Parvalbumins from cold-adapted teleosts are more flexible, with increased reliance on entropy changes associated with binding (30). Residue variation in teleost TnC1a has been suggested to increase molecular flexibility facilitating the protein to be activated by lower concentrations of Ca\textsuperscript{2+} (45). As well as TnC, in fish parvalbumin is another example of a non-catalytic protein following similar patterns of thermal adaptation to those in
enzymes (63), where Ca\(^{2+}\) binding properties vary by species and the environmental temperature of native habitat (31). Orthologs from species adapted to disparate temperatures can have variation in intrinsic stability, which alters the kinetics of enzyme function.

Structural stability is considered to be the primary structural constraint on amino acid substitutions. Stability is created through folding of a protein in a manner that creates the arrangement of residues with the minimum possible free energy, or the minimal level of frustration (130). The structural sub-functionalization process can facilitate the emergence of novel or modified functions through the retention of neutral sequence mutations while still maintaining protein stability (112, 114). Mutations that become fixed under purifying selection such as those in Tn are nonrandom and are typically linked to residue changes that promote stability. TnC is highly conserved with very few residue substitutions that occur in the core of the molecule. Increased structural stability of TnC itself would decrease the affinity for Ca\(^{2+}\) due to an increased energy barrier to change conformation from the apo state (43). TnC1a does have a higher melting temperature than TnC1b, but comparable shifts between the melt temperatures found in the apo and Ca\(^{2+}\) bound states. Changes in the stability of the protein appear to be relatively small in comparison to variations in the conformational flexibility of the protein in the transition between open and closed conformations.

Proteins must be sufficiently stable to hold the proper fold but local instability is often required for function (36). Tn is an example of a highly mobile protein at physiological temperatures (84). Thermal stability relates to the need for the entire enzyme molecule to maintain the appropriate conformational flexibility at physiological temperatures, which allows the conformational changes that accompany binding and catalysis to occur at appropriate rates (119). High stability correlates with compact structures with increased tolerance to mutations (8). Binding proteins or enzymes must retain the ability to undergo conformational changes while being rigid enough to allow the binding of substrates (53). Sequence substitutions that lead to increased flexibility in regions of the enzyme involved in catalysis may reduce energy (enthalpy) barriers to rate-governing shifts in conformation. This can be quantified using methods that localize frustration in proteins, which provides insight into the functional constraints on the evolution of protein energy landscapes (35).
In TnC, conformational changes involving variation between open/closed changes in energetic state may occur separately from Ca$^{2+}$ binding. The regulatory N-TnC fluctuates between open and closed conformation. The shift between open and closed TnC conformations facilitates the likelihood of Ca$^{2+}$ binding. The conformational change, which exposes a large hydrophobic patch is entropically unfavourable and offsets the negative free energy of Ca$^{2+}$ binding (43). This transition in zebrafish TnC1a appears to be more favourable relative to zebrafish TnC1b, suggestive of conformational flexibility between the two states a large determinant of Ca$^{2+}$ sensitivity. Selection pressures thus exist for dissociation kinetics to be just fast enough for transient Ca$^{2+}$ binding rather than as slow as possible, as would be expected in a very stable complex (53).

Accounting for stability has been shown to improve the fit of many models used for phylogenetic analysis (7, 101). Many structural/thermodynamic models incorporated with evolutionary divergence define static interactions of amino acids within a structure without accounting for molecular flexibility or structural optimization. Based on the functionality of TnC, which requires stability and flexibility, neglecting flexibility will not provide high accuracy in detecting sub-functionalization. Functional characterization of Ca$^{2+}$ binding affinities for the physiological impact of residue variations between zebrafish paralogs was determined both through simulations and biochemical measurements. ITC measurements represent a measurement of the equilibrium energy that includes the change in proportion of open and closed TnC molecules as a function of Ca$^{2+}$ titration. The energetics of the shift in equilibrium between open and closed states cannot be experimentally decoupled from the Ca$^{2+}$ binding interaction so the conformational change itself may produce more energy than the Ca$^{2+}$ binding event. Hence this measure may in fact be a combination of structural stability, inter-helical flexibility as well as Ca$^{2+}$ affinity. Our measurements looking only at TnC in Chapter 6 may not be sufficient to fully describe functional variation between paralogs. Highly coevolving residues are often found in flexible regions and facilitate structural transitions by forming and disrupting interactions cooperatively (73). To fully understand the dynamics between the open and closed state of TnC, TnI is necessary to stabilize the open conformation, which results in higher sensitivity. An understanding of the interactions with other proteins is necessary for full understanding of variability in function of Tn rather than limiting the analysis to measurements of Ca$^{2+}$ sensitivity of TnC alone.
8.3. Aim 3: Correlated evolutionary patterns in the interacting members of the Tn complex TnC and TnI

8.3.1. Protein-protein interactions

The function of the Tn complex is intrinsically linked to the evolution of intermolecular interactions that guide sarcomeric function. In Chapter 6, we combined phylogenetic analysis with molecular dynamics simulation to identify divergent residues in the TnI switch region and model their structural and functional effect. Not all functionally divergent sites represented critical sites in the interface of TnI switch region and TnC, but were suspected to have roles in other protein interactions. Structure and evolution are often studied separately, which leaves gaps in our understanding of the intermolecular co-evolution of interacting proteins in multimeric complexes. The evolution of TnC and other proteins constrained by interacting subunits is influenced by factors than simply structural stability of the individual subunits (3). For example, in cytochrome oxidase, the mitochondrial example of intra-genomic coadaptation, higher conservation observed among mtDNA-encoded residues was likely due to factors other than structural stability (3). When the active conformation of a protein is destabilized by a sequence substitution, fitness of the organism can be reduced due to loss of function and the likelihood of that mutation being retained is reduced. The stress model for phylogenetic analysis incorporates stability with the ability of a protein to adopt the correct conformation (67) in an attempt for phylogenetic analysis to more accurately account for protein structure. Appropriate measures of biophysical parameters constrain evolutionary rates of specific residues and thus can be used to improve the robustness of phylogenetic analysis. However, the incorporation of specific structural models in current evolutionary tests are currently limited to intramolecular forces within a single protein (28). The conformation of proteins may be further be stabilized by interacting partners, and their evolution may in fact be constrained by interactions.

On an evolutionary scale, the dosage-balance hypothesis speculates that many retained duplicate genes often encode proteins that participate in a greater than average number of protein-protein interactions, often as members of complexes (58). Interaction
sites that mediate important functions by binding regulatory proteins have strong evolutionary constraints on amino acid substitutions (57). These interaction sites cannot be understood at the level of an isolated protein. On an evolutionary scale, multi-protein complexes are favoured over larger proteins (96). Large proteins are difficult to fold and expensive to synthesize relative to smaller interacting proteins. Individual interactions between proteins can also result in finer control over function than that of an individual protein. Interacting proteins stabilize their binding partners, which acts as a secondary evolutionary advantage. Interacting partners can alleviate selective pressure for stable folding in proteins that are members of a multimeric complex and allow greater conformational flexibility for their function (25). Hence the Tn complex has likely evolved to support the relatively small protein TnC, whose EF-hand motifs play a critical role in Ca$^{2+}$ handling.

Sequence modifications in TnC can drive modifications in TnI. The hydrophobic patch of TnC increases opening increases upon interaction with the switch region of TnI in mammalian models (40, 91). TnI switch peptide binding to TnC stabilizes the Ca$^{2+}$ bound open form of TnC, thus increasing the Ca$^{2+}$ affinity of N-TnC. By extending the amount of time that TnC in this open state, actin-myosin cross bridges may remain intact for longer periods. This accounts for the change in function that results when either TnC paralog is complexed with the switch region of the TnI1.1 paralog instead of the TnI1.5 paralog. Modification of the function of Tn can encompass a range of protein interactions that are not limited to changes in TnC Ca$^{2+}$ affinity. In Ca$^{2+}$ binding proteins, clusters of highly frustrated interactions are necessary for proper flexibility of the protein, which allows for protein–protein assembly and recognition (124). In TnC, functionally divergent residues vary in frustration level and thus interaction likelihood and strength as demonstrated in Chapter 6. Without accounting for interacting partners like TnI, the functional divergence of TnC may not be realized.

The combination of paralogs can influence the functional nature of the interactions in the Tn complex. For example, in vitro cTnT and ssTnI together create less relaxation of cardiac fibers through reduced inhibition of force development in mammals (49). The strength of interaction between the TnC hydrophobic patch and the TnI switch peptide in mammalian N-TnC1 and TnI3 is much less than that seen between N-TnC2 and the
corresponding TnI peptide (35). As the TnC- TnI interaction is responsible for pulling TnI away from its inhibitory position on the actin thin filament, the lower strength of interaction between the cardiac form of TnC/TnI during Ca\(^{2+}\) activation may allow for greater interaction of TnI with actin to further enable the exposure of myosin binding sites. The lower strength of interaction between mammalian cardiac TnC/TnI is responsible, at least in part, for the dependence of cardiac contractility on the formation of strong cross-bridges for full activation to occur. The increased dependence of strong cross-bridge formation has been suggested to be partially responsible for the steeper length dependence of force generation. Variation in teleosts resulting from this interaction may be chamber specific. A weaker interaction between the switch region of TnI1.5 and TnC1a only appears in the ventricle of zebrafish that have been acclimated to warmer temperatures. This could facilitate faster mechanical relaxation, demonstrating how variation in this specific interaction could be important to increased heart rate observed with increased temperature.

A higher degree of residue conservation is essential for residues near to the active site of a protein (15). Functional or structural pressures act to maintain specific subsets of residues at active-site positions within a specific protein, or between interacting proteins. Evolutionary constraints on protein interfaces act to maintain favourable and avoid unfavourable interactions. According to the energy landscape theory, interactions that are present in the native state are energetically stabilized over any possible non-native interactions (71). Framing functional divergence in this way is important for understanding folding patterns, especially when coupled with information from crystallography, as well as being extended to understanding function that may involve multiple conformations. In TnC/TnI interaction, this is found as highly purifying selection on residues that are directly involved in interactions between the switch peptide and TnC. However, due to the high degree of conservation of the entire TnC protein, it is difficult to detect the specific residues involved in the interaction in this manner. Due to the dynamic nature of The TnC/TnI interaction, multiple residues may be playing an important but transient role and detection of divergent residues may identify many compensatory residues that maintain function in the native state. Rearrangement of interfacing residues are a key component of protein conformational changes and can be encoded as coevolving residue pairs (73).
In Chapter 7, the interaction between N-TnC and the TnI switch region was the primary focus. This is not the only domain-specific interaction that places constraints on evolutionary pressure. The majority of divergent residues in TnC occur in the first 12 residues of the N-helix. This region does not interact with the switch peptide and did not factor into changing TnC/TnI interaction strength. The N-helix can play a role in the modulation of Ca2+ sensitivity due to intramolecular interactions that influence the structural stability of site II of TnC (14, 121). Two amino acids in this region, residues 2 and 4 were identified as sites of divergent selection and Type II functional divergence by PAML and DIVERGE, respectively. The N-helix of TnC in mammals has been shown to interact with helix 1 of TnI (95). This region of TnI contains a putative PKC target that is conserved across all vertebrate TnI1/TnI3 paralogs. In mammals, protein kinase C (PKC) targets cTnI S43/45 for phosphorylation. This phosphorylation event plays a role in prolonging twitch duration, possibly by slowing cross-bridge kinetics and decreasing myofilament tension (12, 85, 87, 110). The proximity of residue 2 in TnC suggests a role in propagating the signal from this phosphorylation event. This could explain the importance of the Asn/Asp substitution at this position but further investigation is necessary.

Consideration of co-evolving residues is useful in protein structure prediction. The use of well-characterized structural models to generate evolutionary models strengthens their predictive power. Covariation is observed as pairs of residues that co-occur in multiple sequence alignments more frequently than expected (122). In multimeric protein complexes, covariation occurs when functional or structural pressures maintain specific residues in interacting positions. The detection of selection pressures that facilitate functional divergence at the interfaces of protein-protein complexes can be applied to understanding the evolution of multimeric proteins. In order to optimize interactions at protein-protein interfaces, distinct evolutionary rates for interacting residues are driven by selection for increased stability of the complex (3). X-ray crystallography derived structures represent a single protein conformation but proteins fluctuate among many conformations (101). Consideration of an equilibrium of alternative structures rather than a single static structure is necessary to understand many aspects of protein function. Predictions that are made based on homology structures may be sufficient for broad
evolutionary tests, but for well-characterized proteins like TnC functional divergence can be classified with more detail by considering the dynamic nature of the structures.

Appropriate classification that includes the open and closed states when describing changes in Ca$^{2+}$ handling. When considering Ca$^{2+}$ sensitivity of TnC, we find that conformational changes account for the majority of the variation between paralogs. This suggests that factors influencing the likelihood of the transition between the open and closed state and stabilizing one conformation over another are critical in determining whether a mutation constitutes functional divergence. For TnC, the importance of conformational flexibility may be more important than stability for function especially with the addition of interacting partners that stabilize conformational changes. The power of an integrative predictive approach to problems that are too complex to test may be crucial to understanding the nature of variance across species. The Tn complex examined in this dissertation is one such problem, it is prohibitive time and resource expensive to experimentally test all of possible paralog combinations or sequence substitutions. Our novel approach in Chapter 7 combined traditional use of detection of selection pressure across genomic sequences with detection of functional divergence across protein sequence of paralogs (55) and predictions about divergence of protein structures through molecular dynamics simulation and homology modeling. This integrative approach to functional divergence relies on signals of both regulatory and structural sub-functionalization, and considers evolutionary pressures on the Tn complex as a whole.

8.3.2. The Tn complex as a model of coevolution between subunits

When coevolution is assumed to be present between sites, the sequences are expected to share a common ancestor, and the observed changes to be caused by correlated evolutionary processes. Correlated mutations suggest that compensatory changes that occur between interacting residues are crucial to maintain protein function (108). Based on protein sequence similarity, TnT and TnI have a common evolutionary origin (11) while TnC has an evolutionary origin in common with other Ca$^{2+}$ binding proteins (18, 46). Genes encoding TnI and TnT are closely linked in vertebrate genomes (66). The divergence of muscle-specific isoforms within a single species is greater than
the divergence of each isoform in different species (16). The conservation of a TnI-like epitope structure in Tnl and TnT suggest that the ancestor of both Tnl and TnT was likely a TnI-like inhibitory protein (16). Divergence in muscle-specific gene expression profiles can occur based on cis-regulatory changes even without substantial changes in gene structure relative to the common ancestor (4, 21, 24, 39, 68, 97, 132). The undifferentiated utilization of the same TnC isoform in cardiac and slow skeletal muscles of mammals demonstrates that the duplication of cardiac and slow skeletal TnI-TnT pairs was a relatively recent event relative to the divergence of the fast skeletal paralogs (16).

The three members of the Tn complex must share selective pressures since intramolecular and intermolecular interactions must evolve together in order to maintain structure and function. The function of the Tn complex is far more complicated than the Ca\(^{2+}\) affinity of the regulatory site II of TnC. The conformational changes and stability of TnC are greatly influenced by other binding partners in the thin filament such as Tnl, TnT and tropomyosin. Functionally divergent residues can be used to predict points of interest in changing function such as those that effect conformational change. Considering either structure or function in isolation leads to oversimplification, which limits predictions of functional divergence. For example, hydrophobicity as an index would be greatly influenced by the transition of TnC between the open and closed states (43), as well as the interaction with the Tnl switch (40, 91). The overall stability of TnC is also greatly influenced by interactions with the IT arm of the Tn complex(43). The stabilization of the Tn complex can mediate the conformational flexibility in TnC and change the overall position of TnT in a way that modifies the likelihood of contraction. The evolution of Tnl/TnT as a stabilizing factor or signal transducing adaptor couples the Ca\(^{2+}\) binding of TnC to the actin-myosin contractile machinery (65). The key role in propagating the Ca\(^{2+}\) signal that leads to mechanical contraction necessitates shared selective pressures in order to maintain function.

The troponin complex is well characterized in terms of structure and function, but prior to the work presented here, was unexplored as a model of co-evolution between specific interacting domains. Slowly evolving genes tend to persist because they give sub- or neo-functionalization more time to take place before deleterious mutations occur (113). Even with highly conserved proteins, neutral substitutions are often accompanied by
compensatory substitutions that manifest as co-evolved residue pairs (109). The process of switching between multiple paralogs to produce functional divergence has been established in mammals, but the presence of over-represented duplicates in fish lends itself to a greater capacity for functional through changes in paralog usage. The variation in interaction strength of the switch region of TnI paralogs with TnC described in Chapter 7 provides one example of a consequence of substitutions. While this may not result in drastic changes, compensatory substitutions to prevent loss of function provide an important capacity to adapt to environmental change without losing the ability to cope with sudden changes.

Our example of divergent selection patterns between TnC and TnI paralogs provides the evolutionary basis for variation in the calculated interaction strength between TnC and TnI. This integrates our understanding of both regulatory and structural sub-functionalization in a multimeric protein complex. In addition, limited structural variation in the divergent regions of interacting proteins may still provide large variation in predicted interaction strength, demonstrating that studying subunits of the complex independently is insufficient for understanding functional changes that accompany residue switches. Current models exist of co-evolution between proteins for well characterized structural motifs such as those found in the globin family (54, 88, 89) as well as between subunits of the oxidative phosphorylation protein cytochrome oxidase (22, 107). Co-evolution models are important in describing the evolution of interacting subunits and the coordination of genes from separate genomes in the case of cytochrome oxidase. The use of a well-characterized structural model allows us to investigate the effects of divergence between paralogs more effectively than phylogenetics alone, and the use of evolutionary selective pressures allows prediction of important residues that produce variation in the interaction. This is an important issue in the protein evolution field.

Tn provides an example of a well-characterized complex with three subunits, coded for by genes with distinct evolution history that share selection pressures. The interacting proteins of the Tn complex must also facilitate the balance between conformational position, rigidity and flexibility and ultimately modulate mechanical contraction in response to the binding of Ca^{2+}. Since thermal tolerance is linked to cardiac scope, understanding the functional divergence of a critical regulator of cardiac contraction
will lend understanding to the capacity for variation in ectotherms. The co-evolution of Tn subunits can thus be used for the model of variation in contractile function with multi-faceted dynamic interactions.

8.4. Future Directions

This dissertation approaches Tn as a model of the co-evolution of interacting proteins. While we begin to characterize the divergence of TnC in zebrafish cardiac tissue, further studies are necessary to truly understand the sub-functionalization of paralogs. The nature of regulatory sub-functionalization should be characterized beyond just expression patterns. This involves analysis of promoter/enhancers of TnC to determine if there is divergence in regulation between TnC paralogs and/or other Ca2+ binding proteins. As well, information can be gathered on the conservation of regulation based on motif similarities between Tn sub-units. Further information regarding the co-evolution of Tn subunits may also be revealed as conserved in syntenic blocks across teleost phylogeny. For example, the conservation of localization of Tnl/TnT genes has been well established (65), but linkage to TnC paralogs has not. Regulatory sub-functionalization should also be examined on the post-translational level in teleost Tn. Splice variation plays a role especially in TnT for mammals, but it is less clear how this influences Tn complex composition in fish. The further regulation of function by post-translational modifications such as phosphorylation should be examined. This is especially relevant to Tnl paralog usage and may influence interactions between TnC and Tnl. Each of the sites identified in this thesis are putative phosphorylation sites in Tnl based on mammalian homology with no prior functional characterization in fish.

The changes in regulatory sub-functionalization on the transcript level that are described in this dissertation should also be confirmed on the protein level. Proteomics studies will reveal whether the changes that are predicted by gene expression translate to protein expression. While the presence/absence of transcript is correlated to the presence/absence of the Tnl protein in rainbow trout (2), there is a lack of information regarding accurate quantification and/or turnover rates for the protein products of Tn
paralogs in teleosts. The use of traditional Western blotting techniques is confounded by non-specific antibodies for fish, and often does not have sufficient sensitivity to distinguish between highly conserved paralogs. Advances in proteomics techniques such as mass spectroscopy based multiple reaction monitoring would allow for accurate quantitation of each paralog (29) regardless of level of conservation.

Further work into understanding the nature of structural sub-functionalization can be done for the paralogs that we have already examined. Evidence of conformational changes can be further clarified to determine whether the energy of the conformational change in TnC is actually responsible for Ca\(^{2+}\) affinity differences determined in ITC. Longer time-scale simulations of both the apo and Ca\(^{2+}\) bound state of each TnC paralog will be needed to determine the energy of these conformational changes. This must be examined further in the context of the Tn complex, where interacting partners may provide further stabilization of a particular state, changing the energy threshold of Ca\(^{2+}\) activation of the myofilament. The Tn model can also be used in this context to examine the evolution of local vs global stability in EF-hand proteins. The results in Tn can be compared to Ca\(^{2+}\) binding proteins that lack dedicated adaptors such as TnI/TnT in complex. By utilizing the frustration index across paralogs of TnC protein, predictions about regions that participate in in the stability of interactions can be made (124). Finally, our MD-based predictions of the structure and function of zebrafish TnC can be verified experimentally through NMR.

Our work only begins to look at the interactions within the Tn complex. Further interactions between TnC and other regions of TnI could be examined to explain functional divergence in the N-helix of TnC. Interactions between TnT and the Tn complex have not been considered on a per-residue basis that uses a phylogenetic prediction approach with either TnI or TnC. This would detect whether selection pressures and functionally divergent residue tests hold with those described for other regions of Tn subunits.

While several of our studies speculate on the functional relevance of TnC changes between paralogs, variation within isolated TnC may not be sufficient for physiological relevance. We suspect the changes observed in Ca\(^{2+}\) affinity of TnC paralogs from ITC studies change with increasing levels of complexity, with the addition of the Tn complex, thin filament, and in cardiomyocyte preparations (90). Measurements at multiple levels
would be more informative regarding the functional relevance of Tn paralog combinations, though experimentally far more time and resource intensive than the predictive approach undertaken in this thesis. This could include characterization of Tn complex in isolation, in the thin filament, as well as in cardiomyocyte preparations.

Finally predictions of changes in contractility with paralog switches could be examined using TnC mutant zebrafish. CRISPR/Cas9 can be used to mutate zebrafish embryos so that they only express one paralog or the other. Based on previous studies using TnC1 morpholinos in embryos, the knockdown of one gene or the other should still allow zebrafish to maintain cardiac function and survive to adulthood (118). With exclusive expression of particular Tn paralog combinations, cardiac functional parameters such as EF and E/A could be examined utilizing a similar echocardiographic approach as the one developed here. The variation of specific TnC1 paralogs would experimentally generate the Tn combinations predicted in this thesis in complex for specific chambers of the heart at given temperatures. This could be extended to isolate single cardiomyocytes from fish that have been assessed echocardiographically to check for variation in force of contraction.
8.5. References


11. Brunet NM, Chase PB, Mihajlovic G, and Schoffstall B. Ca\textsuperscript{2+}-regulatory function of the inhibitory peptide region of cardiac troponin I is aided by the C-terminus of cardiac troponin T: Effects of familial hypertrophic cardiomyopathy mutations cTnI R145G and cTnT R278C, alone and in combination, on filament sliding. *Arch Biochem Biophys* 2014.


