Pathogenic peptide deviations support a model of adaptive evolution of chordate cardiac performance by troponin mutations

Nathan J. Palpant,1 Evelyne M. Houang,2 Wayne Delport,3 Kenneth E. M. Hastings,4 Alexey V. Onufriev,5 Yuk Y. Sham,2 and Joseph M. Metzger1

1Department of Integrative Biology and Physiology and 2Center for Drug Design, University of Minnesota Academic Health Center, Minneapolis, Minnesota; 3Antiviral Research Center, Department of Medicine, University of California, San Diego, California; 4Montreal Neurological Institute and Department of Biology, McGill University, Montreal, Quebec, Canada; and 5Departments of Computer Science and Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia

Submitted 16 February 2010; accepted in final form 23 April 2010

Evidentiary pressures favor mutations that mitigate the susceptibility to disease, especially as it pertains to the reproductive fitness of an organism (5, 48). Consequently, monogenic diseases provide excellent reference points from which to understand the implications of certain mutations as a retrospective view on evolutionary modifications (5, 48). In other words, human disease resulting from pathogenic peptide deviations, defined here as single nucleotide polymorphisms (SNPs) that cause pathologic recapitulations nonmammalian physiology, provide insight into the influence of selective pressures on specific residues or domains of a given protein to accommodate highly specified physiological requirements (5, 48, 54). This study assessed phylogenetic changes that occurred during evolution of mammalian myofilament calcium sensitivity with a focus on troponin, the regulatory molecular switch complex of the myofilaments.

Troponin is a heterotrimeric protein complex comprised of three subunits. These include the tropomyosin (Tm) binding subunit, troponin T (TnT), the calcium binding subunit, troponin C (TnC), and the actomyosin ATPase inhibitory subunit, troponin I (TnI) (Fig. 1A). The coordinated actions of the troponin complex are specifically designed to regulate the functions of actomyosin cross bridges in a calcium-dependent manner during the rhythmic transitions between contraction (systole) and relaxation (diastole) (38).

During systole, myofilament activation is mediated through binding of the TnI switch arm to an exposed hydrophobic patch on the calcium-bound NH2-terminal domain of cardiac (c)TnC (4, 41, 45). These actions enable coordination of the azimuthal positioning of Tm in the actin groove by TnT and subsequent sarcomeric contraction mediated by actin-myosin cross-bridge cycling (38, 66, 100). The biophysics of myofilament contraction are translated into ventricular inotropic performance at the whole organ level (38). During diastole, calcium is released from cTnC and sequestered into the sarcoplasmic reticulum, allowing the regulatory arm of TnI to translocate into an inhibitory position on actin. This is sufficient to inhibit actin-myosin cross-bridge activation resulting in sarcomeric relaxation (60, 71). Förster resonance energy transfer (FRET) experiments have shown that the movements of the regulatory arm of TnI (TnIreg) during systole and diastole are essential for the transmission of the calcium signal to the rest of the myofilament proteins (60, 71).

The specific amino acid composition of the cTn complex markedly affects its sensitivity to activating calcium, which, ultimately, influences contraction (inotropic) and relaxation (lusitropic) performance (20, 21, 33, 37, 92, 112). Compared with mammals, studies of several species suggest that ectotherms such as fish and amphibians require heightened myofilament calcium sensitivity to maintain cardiac performance in the context of challenging environmental conditions such as cold temperatures and extended periods of hypoxia (32, 34, 35, 61, 62, 88, 89). However, SNPs in cTn that cause increased myofilament sensitivity to activating calcium in mammals, termed here pathogenic peptide deviations, result in cardiac disease including arrhythmias, diastolic dysfunction, and increased susceptibility to sudden cardiac death (9, 20, 72, 101). These clinical observations reveal an absolute requirement for reduced myofilament sensitivity to activating calcium within...
the mammalian taxa compared with other phylogenies. Using large-scale bioinformatics and an array of molecular modeling techniques, we propose that the intrinsic relaxation potential of cTn has been enhanced during chordate evolution to meet the functional requirements of the mammalian cardiovascular system.

METHODS

Amino acid and mRNA sequences and alignments. Sequences were extracted by search of NCBI protein and nucleotide databases as well as the UniProt KB protein database. Defining the taxonomic relationship between species studied was accomplished by use of the NCBI taxonomy database (Supplemental Fig. S1). Table 1 outlines the abbreviations used to reference specific species. Supplemental Table S1 outlines all relevant details on the sequences acquired from NCBI or UniProt KB databases used in this phylogenetic analysis. To be consistent with prior studies, all amino acid numbering in this study refers to the adult rat isoform sequence (cTnI, cTnC, and cTnT) including the starting methionine unless otherwise stated. Protein and nucleotide sequences were aligned by ClustalW analysis as originally described by Thompson et al. (99). Multiple alignment parameters were as follows: gap penalty = 15; gap length penalty = 6.66; protein weight matrix: Gonnet series. Pairwise analysis of percent identity between sequences was used to define the relationship between proteins or within proteins across designated species.

Genome scans for predicted TnI genes. Sequences for TnI isoforms were acquired by whole genome scans of numerous species. Genome BLAST analysis for TnI isoforms was performed on Anolis carolinensis (lizard, UCSC), Taeniopygia guttata (zebra finch, NCBI), Cavia porcellus (guinea pig, UCSC), and Callithrix jacchus (marmoset, UCSC). Query sequences from taxonomically related species were used for the genome scans. Predicted isoform for TnI sequences was accomplished by ClustalW analysis followed by sequence percent identity and phylogenetic tree analysis. The Kimura distance formula was used to calculate distance values, derived from the number of nongap mismatches and corrected for silent substitutions. The values computed are the mean number of differences per site and fall between 0 and 1. Zero represents complete identity and 1 no identity. Scale of the phylogenetic tree indicates the number of nucleotide substitutions per 100 residues. Phylogenetic tree analysis was used to confirm identity of predicted TnI isoforms by sequence correlation with other TnI isoforms. Sequence ID and divergence values were calculated to define the relationship between novel TnI proteins and related TnI isoforms to assist in confirmation of postulated isoform designation. Divergence(i,j) is calculated as {100[Distance (i,j)/Total Distance]}, where (i,j) is the sum of the branch lengths between two sequences and Total Distance is the sum of branch lengths. All data regarding protein and mRNA sequences are outlined in Supplemental Table S2 and Supplemental Figs. S2 (nucleotide) and S3 (protein). Determination of evolutionary selective pressure. Phylogenetic codon models were used to understand the selective pressure both on the full gene sequence of cardiac TnI and on the switch domain. Briefly, codon-based methods (recently reviewed in Refs. 2, 24) make use of a continuous-time Markov process to model synonymous (dN) and nonsynonymous (dS) substitutions along a phylogenetic tree. Phylogenetic trees were estimated, with PHYML (40), independently for the full gene sequence and for the switch domain, since only a subset of full cTnI sequences were available for the mammalian taxa of interest. The switch domain is short (63 nucleotides), and thus some internal branches of the phylogeny had lengths of zero; we collapsed those branches into polytomies. It was assumed that synonymous and nonsynonymous substitution rates at codon sites can be approximated by using discrete distributions the parameters of which are estimated by maximum likelihood. To rule out the potential confounding effect of recombination or gene conversion, the alignment was screened with GARD (51, 52); no evidence for recombination was detected. To identify the distribution that best describes synonymous and nonsynonymous rate variation across all sites, a stepwise procedure was used starting at a single rate for all sites and progressively adding new rate classes, each time estimating the synonymous and nonsynonymous rates of each class and evaluating model fit with the small sample correction to Akaike’s information criterion (AIC-c) (93). The stepwise iteration was terminated once no further improvement in model fit was achieved. A fixed effects likelihood (FEL) method (implemented on www.datamonkey. org; Ref. 80) was used to estimate the synonymous to nonsynonymous substitution rate at each site independently for the switch domain. Briefly, evidence for either positive (dN/dS > 1) or purifying
equal to that in which this constraint is relaxed. Statistical significance in which synonymous and nonsynonymous rates are constrained to be superpositioning the two structurally conserved Ca$_2^+$ comparison between cTn and skeletal (s)Tn was carried out by 2005 forcefield to optimize all hydrogen bonding networks. Structure

Table 1. Species taxonomy and abbreviations

<table>
<thead>
<tr>
<th>Mammals</th>
<th>Homo sapiens</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>Pongo abelii</td>
<td>SO</td>
</tr>
<tr>
<td>Sumatran orangutan</td>
<td>Macaca mulatta</td>
<td>Rh</td>
</tr>
<tr>
<td>Chimpanzee</td>
<td>Pan troglodytes</td>
<td>Ch</td>
</tr>
<tr>
<td>Rhesus</td>
<td>Callithrix jacchus</td>
<td>Ma</td>
</tr>
<tr>
<td>Marmoset</td>
<td>Bos taurus</td>
<td>B</td>
</tr>
<tr>
<td>Bovine</td>
<td>Equus caballus</td>
<td>Ho</td>
</tr>
<tr>
<td>Horse</td>
<td>Capra hircus</td>
<td>G</td>
</tr>
<tr>
<td>Goat</td>
<td>Ovis aries</td>
<td>Sh</td>
</tr>
<tr>
<td>Sheep</td>
<td>Sus scrofa</td>
<td>P</td>
</tr>
<tr>
<td>Pig</td>
<td>Canis lupus familiaris</td>
<td>D</td>
</tr>
<tr>
<td>Dog</td>
<td>Felis catus</td>
<td>C</td>
</tr>
<tr>
<td>Cat</td>
<td>Cavia porcellus</td>
<td>GP</td>
</tr>
<tr>
<td>Guinea pig</td>
<td>Orcytolagus cuniculus</td>
<td>Rab</td>
</tr>
<tr>
<td>Rabbit</td>
<td>Rattus norvegicus</td>
<td>R</td>
</tr>
<tr>
<td>Rat</td>
<td>Mus musculus</td>
<td>M</td>
</tr>
<tr>
<td>Mouse</td>
<td>Monodelphis domestica</td>
<td>O</td>
</tr>
<tr>
<td>Opossum</td>
<td>Ornithorhynchus anatinus</td>
<td>Pi</td>
</tr>
<tr>
<td>Platypus</td>
<td>Anolis carolinensis</td>
<td>L</td>
</tr>
<tr>
<td>Lizard</td>
<td>Gallus gallus</td>
<td>Ch</td>
</tr>
<tr>
<td>Birds</td>
<td>Coturnix coturnix</td>
<td>Q</td>
</tr>
<tr>
<td>Chicken</td>
<td>Meleagris gallopavo</td>
<td>T</td>
</tr>
<tr>
<td>Quail</td>
<td>Taeniopygia guttata</td>
<td>ZF</td>
</tr>
<tr>
<td>Turkey</td>
<td>Xenopus laevis</td>
<td>X</td>
</tr>
<tr>
<td>Zebra finch</td>
<td>Xenopus (silurana) tropicalis</td>
<td>Xt</td>
</tr>
<tr>
<td>Amphibians</td>
<td>Rana catesbeiana</td>
<td>BF</td>
</tr>
<tr>
<td>Western clawed frog</td>
<td>Bufo marinus</td>
<td>MT</td>
</tr>
<tr>
<td>Fish</td>
<td>Salmo salar</td>
<td>S</td>
</tr>
<tr>
<td>Salmon</td>
<td>Danio rerio</td>
<td>Z</td>
</tr>
<tr>
<td>Zebrafish</td>
<td>Tetradon fluvialis</td>
<td>GP</td>
</tr>
<tr>
<td>Green puffer fish</td>
<td>Polysteus senegalus</td>
<td>DE</td>
</tr>
<tr>
<td>Dinosaur eel</td>
<td>Lehtenetro japonicum</td>
<td>AL</td>
</tr>
<tr>
<td>Arctic lamprey</td>
<td>Ascidinai 1</td>
<td>As1</td>
</tr>
<tr>
<td>Invertebrates</td>
<td>Ascidinai 2</td>
<td>As2</td>
</tr>
<tr>
<td>Ascidian</td>
<td>Halocynthia roretzi</td>
<td>A</td>
</tr>
<tr>
<td>Amphioxias</td>
<td>Branchiostoma belcheri</td>
<td>A</td>
</tr>
</tbody>
</table>

(dN/dS < 1) selection was identified by comparing the fit of a model in which synonymous and nonsynonymous rates are constrained to be equal to that in which this constraint is relaxed. Statistical significance was evaluated with a likelihood ratio test (P < 0.05).

Molecular modeling and structural analysis. All modeling was carried out with Maestro version 9.0.111 (Schrödinger, New York, NY; 2007). The solved X-ray crystallographic structures of human cardiac troponin (95) (PDB: 1J1E, 3.3 Å) and skeletal troponin (103) (PDB: 1YTZ, 3.0 Å) in complex with Ca$^{2+}$-bound TnC were taken from the Protein Data Bank. The missing hydrogen atoms were added to both X-ray structures, followed by energy minimization with OPLS 2005 forcefield to optimize all hydrogen bonding networks. Structure comparison between cTn and skeletal (s)Tn was carried out by superpositioning the two structurally conserved Ca$^{2+}$-bound TnC. From this superpositioning, truncated hybrid and homotypic crystal structures were analyzed, including cTnC/cTnI, cTnC/sTnI, cTnC, cTnI, and sTnI. The cTnC-TnI regulatory complex was isolated by exclusion of all other residues in the crystal structure except sTnR115-L140, cTnR148-L173, and each TnI was complexed with cTnC M1-K90.

Molecular dynamics simulation. Molecular dynamics (MD) simulations were performed for sTnI-cTnC, cTnI-cTnC, and sTnIH132A-cTnC with the use of NAMD (78) version 2.6 with CHARMM27 forcefield (63). Computational requirements were acquired by using the University of Minnesota supercomputer BladeCenter Linux cluster. All X-ray structures and derived mutants were prepared by CHARMM version c35b1. Each of the complexes was solvated with an explicit TIP3P water model (47) using a simulation solvent box with a 15-Å buffer region around the protein. Sodium counterions were placed at 5 Å from the box boundary to neutralize the system. The solvated system was initialized with 5,000 steps of conjugate gradient energy minimization with protein heavy atoms restrained at 50 kcal/(mol-Å$^2$). The system was then gradually heated with the same restraint from 25K to 300K at 25K increments at 10- ps intervals for 100 ps, followed by a 100-ps equilibration with gradual removal of the heavy atoms restraint at 10- ps intervals under NVT condition. The final unrestrained equilibration was carried out for 100 ps, followed by 40 ns of production simulation at 1 atm and 300K NPT condition. All simulations were carried out with periodic boundary condition by using particle mesh Ewald (PME) (27) with the SHAKE (83) method employed to constrain bond lengths involving hydrogen atoms. The time step in the simulations was 2 fs, with coordinates saved at 1-ps time intervals, resulting in a total of 40,000 configurations for analysis. Two MD simulations were carried out for each of the modeled complexes by random initialization of the starting velocities.

Interatomic distance calculations. All distances are measured in angstroms over the 40-ns simulations for each of the simulated trajectories. For the wild-type (WT) and H132A simulation of the sTnl-cTnC complex, the distance was evaluated between cTnC E19 and sTnI-cTnC complex, the distance was evaluated between cTnC E19 and sTnI H132 (ND1) or sTnI H132A (CB).

Root mean square fluctuations. The root mean square fluctuations (RMSF) for the Ca atoms of each residue were calculated over each of the 40-ns simulated trajectories by using a CHARMM script to evaluate the relationship

$$\text{RMSF} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (X_i(t) - X_{i,j})^2}$$

where $T$ is the trajectory time length, $X_i(t)$ is the average position of each individual residue $i$ at simulation time $t$, and $X_{i,j}$ is the initial position of each residue. The residue fluctuations for TnC and TnI were evaluated independently of each other by superimposing all snapshots to the X-ray structure of each domain as the reference structure.

Theoretical isoelectric point calculations. In the absence of crystal structures, values for the theoretical isoelectric point (pI) for a given amino acid sequence were calculated with the algorithm from ExPASy’s Compute pI/Mw program (12).

Theoretical pK calculations of Tn. The general continuum electrostatics methodology widely used to calculate the energetics of proton transfer is described elsewhere (7, 8). Briefly, the difference between a side chain’s pK and the pK of the corresponding model compound in free solution is determined by the combined effect of two distinct contributions to the total electrostatic (free) energy change. First is the “Born term” or desolvation penalty, which always penalizes burial of a charge inside a low dielectric medium. Second is the “background term,” which represents the electrostatic interactions of the group in question with all other fixed charges in the molecule not belonging to any titratable groups. These energy terms, as well as the matrix of side-site interactions $W_{ij}$, are estimated through a sequence of finite-difference Poisson-Boltzmann calculations in which sites in the protein and their corresponding model compounds have their charge distributions set to those of the protonated or deprotonated form, and suitable energy differences are taken. The methodology allows for a straightforward decomposition of the total computed pK value as a sum of contributions $\Delta pK_i$ from all other groups in the protein, both titratable and nontitratable. Thus the individual $\Delta pK_i$ values can be used to characterize the strength of electrostatic interactions between neighboring groups in protein.

Here we used a web-based implementation of the model available via the H++ server (39) (http://biophysics.cs.vt.edu/H++/). In this work, the protein is treated as a low dielectric medium $\varepsilon_m = 10$, while...
the surrounding solvent is assigned a high dielectric constant \( \varepsilon_{\text{sol}} = 80 \). Consistent with the premise of the continuum solvent models, all water molecules and ions present in the original PDB structures were removed before the calculation. The electrostatic screening effects of (monovalent) salt enter via the Debye-Huckel screening parameter \( \kappa \), which is set to correspond to physiological conditions of 0.145 M salt concentration. The choice of internal dielectric \( \varepsilon_{\text{m}} = 10 \) is justified for titratable groups that are not buried deeply inside the protein matrix, which is the case for the groups of interest in this study. Further justification for the above choice of \( \varepsilon_{\text{m}} \) was justified by comparison of pK values and protonation states of key groups computed by two completely different methodologies: \( \text{H}^+ \) (continuum electrostatics) and \( \text{PROPKA}(58) \) (empirical). Specifically, the predictions of the large upward shift of HIS130 pK differ by only ~0.2 pK units between the two methods.

**Computational alanine scanning: preparation of Ala mutant structures.** For each amino acid on the sTnI-cTnC binding interface we prepared its alanine substitution mutant by deleting all side chain atoms beyond C\( \beta \) in the corresponding WT structure. pK values of titratable groups in the WT and mutant structures were then calculated as described above, and their protonation states were set accordingly assuming physiological pH (pH = 7.2).

**Computation of changes in binding free energies upon Ala substitutions.** A physical model for protein-protein binding free energies was used (49); the model is based on all-atom description and includes energetic contributions from VdW interactions, solvation, and hydrogen bonding. Here we employ a web-based implementation (Robetta: http://robetta.bakerlab.org/alaninescan) of the model (50) with default parameter settings. The alanine scan was performed individually for each group on the binding interface of both partner proteins (sTnI-cTnC). The methodology computes change \( \Delta G_0 \) in protein-protein binding free energy upon alanine substitution of residue “X” in position “NN”; \( \Delta G_{\text{ala}} = \Delta G_{\text{WT}} - \Delta G_{\text{MUT}} \), where \( \Delta G_{\text{WT}} \) is the protein-protein binding free energy of the WT complex and \( \Delta G_{\text{MUT}} \) is that of the corresponding XNNA mutant (e.g., H132A). Thus, \( \Delta G_{\text{ala}} > 0 \) indicates destabilization of the complex in the mutant form relative to the WT form.

Most of the titratable groups in both binding partner protein structures were previously determined to be in their “standard” protonation states; furthermore, these states did not change upon protein-protein complex formation or Ala substitutions of individual interface groups. For these groups no additional calculations were necessary; the \( \Delta G_0 \) values computed by the Robetta server are reported directly in Supplemental Table S3.

However, a calculation based on the H++ methodology reveals that the protonation state of His132 at pH = 7.2 does change upon formation of the complex, resulting in the need to consider an additional pH-dependent contribution to \( \Delta G_{\text{ala}} \), which we denote as \( \Delta \Delta G_{\text{ala}}(\text{pH}) \): \( \Delta \Delta G_{\text{ala}} = \Delta G_{\text{ala}}^0 + \Delta \Delta G_{\text{ala}}^\text{pH} \). To compute this additional term, we follow earlier works (65, 117) and express the total binding free energy as a sum of intrinsic and pH-dependent parts:

\[
\Delta G = \Delta G^0 + \Delta G_{\text{pH}}
\]

Equation 1

where \( \Delta G^0 \) is the pH-independent (intrinsic) binding free energy computed when the site of interest is in a reference protonation state, and

\[
\Delta G_{\text{pH}} = -RT \ln \left( \frac{1 + 10^{\text{pK}^\text{WT}-\text{pH}}} {1 + 10^{\text{pK}^\text{MUT}-\text{pH}}} \right)
\]

Equation 2

where indices C and F denote the complexed and free states of the two proteins, respectively (65), and R and T are the gas constant and absolute temperature, respectively. We take the reference protonation state of His to be singly protonated and Glu to be deprotonated, which are the “standard” protonation states assumed by the Robetta model. With this assumption, \( \Delta G_{\text{WT}} \) and \( \Delta G_{\text{MUT}} \) computed by the model correspond to pH-independent, intrinsic \( \Delta G^0 \) in Eq. 1. Applying Eq. 2 to both WT and MUT, we obtain the following expression for the pH-dependent part:

\[
\Delta \Delta G_{\text{ala}}(\text{pH}) = -RT \ln \left( \frac{1 + 10^{\text{pK}^\text{WT}-\text{pH}}} {1 + 10^{\text{pK}^\text{MUT}-\text{pH}}} \right)
\]

For His132 and Glu19, for consistency, the values reported in Fig. 4C are \( \Delta \Delta G_{\text{ala}} = \Delta G_{\text{ala}}^0 + \Delta \Delta G_{\text{ala}}^\text{pH} \). For His132 the pH-dependent contribution is comparable to the intrinsic part \( \Delta G_{\text{ala}}^0 \) at the relevant pH = 7.2; for Glu19 this additional contribution is very small since pK of this group is well below 7.2.

**RESULTS AND DISCUSSION**

The principal goal of this study was to assess how cTn proteins have been structurally modified during evolution to accommodate the physiological requirements of mammalian myofilament function. To accomplish this, large-scale phylogenetic analysis was performed by ClustalW alignment of 104 primary amino acid sequences of troponin regulatory components including cTnT, cTnC, slow skeletal (ss)TnI, and cTnI from species throughout chordate evolution (Supplemental Figs. S1–S6, Supplemental Tables S1, S2). Table 1 lists species abbreviations summarizing the relevant taxonomy.

**Disease-causing polymorphisms in COOH terminus of cTnI recapitulate nonmammalian physiology.** The positional relationship between disease-causing SNPs, functional domains, and the secondary structural regions of TnI are outlined in Fig. 1. Benign and disease-causing SNPs have been identified along the full length of the human cTnI locus. Over 94% of known disease-causing SNPs are located between the start of the inhibitory region (IR) and the end of the COOH terminus (Fig. 1B). Collectively, this is referred to as the regulatory domain of TnI (95). One of the physiological outcomes observed in these cardiomyopathies is increased myofilament calcium sensitivity (9, 20, 81, 101). The concentration of disease-causing mutations within the regulatory region of TnI indicates that single amino acid alterations in this critical region of the molecule are not tolerated and have significant deleterious effects on TnI function that cause heart disease.

Studies have suggested that, compared with the physiological constraints of mammals, environmental demands such as cold temperatures and hypoxic conditions require heightened myofilament calcium sensitivity to maintain sufficient cardiac performance in several fish (e.g., salmonids) and amphibian species (13, 32, 34, 35, 61, 62, 88, 89, 92). These observations indicate that heightened myofilament sensitivity to activating calcium is a feature of human cTnI disease-causing SNPs that recapitulates normal, healthy nonmammalian myofilament function.

**Phylogenetic analysis of slow skeletal and cardiac troponin I functional domains.** Human SNPs in TnI that cause pathologies recapitulating nonmammalian physiology provide a retrospective view of evolutionary selective pressures. Accordingly, these SNPs lend insights into the evolutionary basis for observed modifications in cTnI during molecular evolution of the mammalian myofilaments (5, 48). Given the observed intolerance of SNPs in the COOH terminus of TnI (Fig. 1B), we hypothesized that evolutionary pressures tightly regulated changes in these regions to modify the performance of Tn in accordance with the evolution of mammalian requirements for sensitivity to activating calcium. To begin, the
primary amino acid sequence of subdomains in TnI were analyzed to understand the phylogenetic basis for differences within these functional regions of the molecule between chordate species (Fig. 2).

Isoforms of TnI are differentially expressed among chordate species. In amphibians the cTnI isoform has been found to be the only TnI gene expressed at the onset of heart formation in the larva and continuing in the adult heart of anuran species (25, 104, 105). However, in avian and mammalian species the ssTnI gene is transiently expressed during heart development and is replaced in the adult heart by the cTnI gene (77, 84, 91). On the basis of these different expression profiles, we hypothesized that comparative analysis of TnI isoforms expressed in the heart would provide a basis with which to interpret modifications in TnI during evolution of chordate cardiovascular systems.

To discern evolutionary pattern differences among the functional domains of TnI, discrete segments of the molecule were aligned and compared across all species. The domains of TnI analyzed included the NH2 terminus (1–44), NH2-terminal TnC binding domain (45–65), IT arm (91–136), IR (138–149), switch arm (152–172), and COOH terminus (153–210). To determine whether amino acid sequences within these functional domains were different between ssTnI and cTnI across chordate organisms, sequence alignment analysis was performed on species in which both isoforms are known (Fig. 2A). This analysis showed that, in contrast to NH2-terminal domains, regions within TnIreg (residues 138–210) all showed a significant reduction in sequence conservation within the cTnI isoform compared with ssTnI (P < 0.05). Furthermore, the cTnI switch arm had the greatest difference in sequence percent identity compared with the ssTnI switch arm (Δ%ID = −20.1 ± 1.9%; P < 0.05 compared with all other domains).

To understand how this change in switch arm conservation correlates with specific phylogenetic groupings, the ssTnI:cTnI sequence percent identity was compared in a pairwise fashion within each class (mammalian, avian, amphibian) (Fig. 2B). This analysis showed that among amphibians the switch arm of ssTnI and cTnI isoforms are more similar than is the case for birds. The greatest divergence (lowest %ID) in ssTnI:cTnI sequence conservation was observed in mammals. In contrast, other functional regions such as the NH2-terminal TnC binding domain showed no difference in ssTnI:cTnI sequence conservation among the taxa studied (Fig. 2B).

These phylogenetic data indicate that the switch arm was a site of rapid ssTnI:cTnI sequence divergence most evident within the mammalian lineage. A vast literature supports the conclusion that the properties of the TnI switch arm can markedly alter the mechanism by which troponin regulates the contractile properties of the myofilament (22, 38, 60, 69, 71, 95, 103, 107, 108, 110–113, 115).

Analysis of amino acid composition of TnI switch arm. We next sought to determine how the mammalian switch peptide was modified. To begin, the theoretical isoelectric point (pI) was calculated for the switch arm fragment for each class (Fig. 3Ai). These data show that the ssTnI switch arm pI did not change across any classes. In contrast, the pI for the cTnI switch arm decreased from amphibian (9.5 ± 0.5) and bird (8.75 ± 0.0) species to mammals (6.76 ± 0.0; P < 0.05 vs. mammalian ssTnI). These data also show no significant difference between the TnI isoform (ssTnI vs. cTnI) pI for amphibian and avian species. Comparatively, there is a significant reduction in the pI for mammalian ssTnI vs. cTnI (9.99 ± 0.0 vs. 6.76 ± 0.0; P < 0.05). These data are supported by crystal structure pI analysis (Supplemental Table S3). Linear regression analysis showed that evolutionary pressure to modify the switch arm peptide sequence between ssTnI and cTnI is significantly related to alterations in the pI of the domain (r2 = 0.84, P = 0.0002) (Fig. 3Aii). Evidence that the ssTnI switch arm sequence composition and pI are not different across classes indicates that this isoform is not being modified during chordate evolution. In contrast, modifications in ionizable residues in the cTnI switch arm are altered during evolutionary peptide adjustments of the mammalian cTn complex.

Sequence analysis of the troponin I switch arm. To further study the observed modification of the switch arm in cTnI from amphibians to avian and mammalian species, predicted protein and mRNA sequences of cTnI from numerous chordate species were collected. In addition, putative TnI sequences were identified by BLAST analysis of genomes from Anolis carolinensis (lizard), Taeniopygia guttata (zebra finch), Cavia porcellus...
(guinea pig), and *Callithrix jacchus* (marmoset) (see Supplementary Table S2 and Supplemental Figs. S2 and S3). Isoform-dependent variations in ionizable residues between ssTnI and cTnI in the switch arm were studied (Fig. 3B). This analysis showed that a nonpolar hydrophobic moiety at position 164 in the switch arm of cTnI is 100% conserved in mammals but was found to be a histidine in all other chordate species and TnI isoforms (with the exception of zebra finch cTnI). Although alanine 164 was identified in all other mammals, a proline residue was surprisingly identified in the predicted platypus sequence. Phylogenetically, platypus, an egg-laying mammal, is situated early in the mammalian lineage (Supplemental Fig. S1). This modification of an amino acid that is perfectly conserved through chordates prior to the bird/mammal divergence and subsequently altered (charged to uncharged or vice versa) only in mammals and thereafter perfectly conserved was not observed at any other locus in the entire troponin complex (cTnI, ssTnI, cTnT, or cTnC). Such a significant level of residue conservation suggests that the transition from a hydrophilic and polar amino acid to a hydrophobic and nonpolar sequence. Physiologically, platypus, an egg-laying mammal, is situated early in the mammalian lineage (Supplemental Fig. S1). This modification of an amino acid that is perfectly conserved through chordates prior to the bird/mammal divergence and subsequently altered (charged to uncharged or vice versa) only in mammals and thereafter perfectly conserved was not observed at any other locus in the entire troponin complex (cTnI, ssTnI, cTnT, or cTnC). Such a significant level of residue conservation suggests that the transition from a hydrophilic and polar amino acid to a hydrophobic and nonpolar sequence.
moiety within the switch arm is likely a key modification in Tn during evolution of the mammalian heart. This is supported by studies showing the importance of this TnI histidine button in myofilament function (22, 23, 75, 76, 112). Analysis of selective pressure at the nucleotide level indicates that the mammalian cTnI switch arm is under significant purifying selection, suggesting that the current peptide composition has long been highly favorable for the proper regulatory functions of troponin in the mammalian heart (Supplemental Table S4).

Crystal structure analysis of calcium-saturated troponin complex. The phylogenetic data indicate that the switch arm of amphibian cTnI is very similar to ssTnI from across all classes in terms of amino acid sequence and pl (Fig. 3). Studies have also shown that cTnI physiology seen in fish and amphibians is functionally recapitulated by ssTnI in the mammalian heart in vivo and in vitro (1, 33, 57, 67, 68, 89, 102, 113, 114). Importantly, what we know about mammalian ssTnI may therefore assist in our understanding of the structure/function relationship of anuran and fish cTnI.

In recent years the crystal structures of the calcium-saturated cTnI complex (95) (PDB: 1J1E, 3.3 Å) and fast skeletal (fs)TnI complex (103) (PDB: 1YTZ, 3.0 Å) have been reported. We assume that the regulatory domain of the putative ssTnI complex mirrors that of the reported fsTnI complex (for justification of this assumption and structural orientation see Supplemental Fig. S7). We propose that comparative molecular modeling of sTnI and cTnI through hybrid [sTnI (103) with cTnC (95)] and homotypic [cTnI with cTnC (95)] atomic resolution crystal structures provides insight into evolutionary modifications that occurred in cTn to accommodate the functional requirements of the mammalian heart.

We find that the structural alignment of the sTnI and cTnI crystal structures in complex with cTnC accentuates the differences in the regulatory interaction between these protein complexes (Fig. 4A). Comparison of the sTnI and cTnI crystal structure shows a marked alteration in the position of TnI helix 4 between sTnI and cTnI relative to helix A and the hydrophobic patch of TnC (Fig. 4A). The structural alignment indicates that H3 has similar positioning in the two complexes. The structure of sTnI indicates that unstructured helix 4 has His132 engaging helix A of TnC in very close proximity to a key acidic residue, Glu19, in the hydrophobic patch of TnC (Fig. 4A). This analysis shows that TnI H132 [protonated, acidic dissociation constant (pKₐ) 8.25] and TnC E19 (pKₐ 2.51) are close enough to develop a salt bridge (Supplemental Table S3; Fig. 4A, inset). This computational analysis indicates that the theoretical binding energy of this interaction is 5.24 kcal/mol, the strongest interatomic bonding observed between any two ionizable groups analyzed in these crystal structures (Fig. 4B). Decomposition of pKₐ data indicates that the pH titration characteristics of H132 and E19 are significantly influenced by the reciprocal electrostatic interactions between these residues (Supplemental Fig. S8, a and b; Supplemental Table S3). The characteristics of this key interaction between cTnC and sTnI (and thus its functional importance) are also observed in another unique histidine-aspartate interaction (H146-D948) in hemoglobin (HbA) that forms a salt bridge and singularly contributes to about half of the maximal alkaline Bohr effect of human HbA (10).

In contrast to the sTn structure, at the H3-H4 linker and into the structured H4 of cTnI, Ala164 translocates away from helix A and the hydrophobic patch of TnC (Fig. 4A). This computational analysis indicates that the regulatory interaction site between the switch arm of TnI and the NH₂-terminal lobe of cTnC: cTnC (red), sTnI (blue), cTnI (yellow). Inset highlights the critical interaction between sTnI H132 and cTnC E19. B: intramolecular interactions between ionizable groups (kcal/mol) in TnI and TnC are shown for the sTnI-cTnC structure (top) and the cTnI-cTnC structure (bottom). Interactions are organized into intramolecular interactions within TnI (left) and cTnC (middle) and intermolecular interactions at the interface between TnI and cTnC (right). The strength of interaction of cTnC E19-sTnI H132 is indicated in orange. Vertical bar = 1 kcal/mol. C: alanine scanning analysis of the change in binding free energy (ΔΔG) among all residues at the sTnI-cTnC regulatory interface. Important contributions to ΔΔG by cTnC E19 and sTnI H132 are shown in orange.

A and the hydrophobic patch of TnC (Fig. 4A). Residue ionization titration analysis confirmed the hypothesis that cTnC and cTnI have limited electrostatic contact at this critical interaction site (Fig. 4B; Supplemental Table S3, Supplemental Fig. S8). Crystal structure analysis showed the strongest electrostatic contacts within the cTnI-cTnC complex to be over four times lower than E19-H132 of cTnC-sTnI.

To gain molecular insights into the marked differences between the sTnI-cTnC and cTnI-cTnC regulatory complexes, computational simulations were carried out on a nanosecond timescale to study the molecular dynamics and stability of the two complexes in atomistic detail. These data show that the switch arm (unstructured H4 in particular) retains strong interfacial contacts between sTnI and cTnC but becomes highly
mobile, molecularly untethered, in the cTnI-cTnC structure [Supplemental Videos S1 (sTnI-cTnC) and S2 (cTnI-cTnC)].

The performance characteristics of the tropinin complex have been shown to be markedly different between myofilaments containing sTnI vs. cTnI (59, 108, 110, 113). Two-dimensional NMR spectroscopy analysis of amide chemical shifts during calcium titration has shown that the switch peptide dissociation constant for cTnI is six times weaker than that of sTnI (59). Transgenic and gene transfer studies have shown that myofilaments containing sTnI have sarcomere dynamics including hypercontractility, slow relaxation, and increased propensity for arrhythmias compared with myofilaments containing cTnI (9, 112, 118). Furthermore, consistent with the concept of a pH-responsive histidine button, TnI proteins with a histidine moiety in the switch arm (anuran cTnI, mammalian ssTnI, or mammalian cTnI A164H) perform as a molecular rheostat capable of maintaining inotropy in the context of severe acidosis or other pathologies (3, 18, 22, 75, 76, 102, 112–114). Myofilaments containing cTnI with an alanine in the switch arm (A164), by contrast, are highly susceptible to acidosis-induced contractile failure. Together, these structural and functional data indicate that significant alterations in the peptide composition and, consequently, tertiary protein structure of the cTn regulatory interaction markedly alter the functional properties of Tn.

We next tested the hypothesis that breaking of the His132-E19 interaction would markedly decrease the thermodynamic affinity between sTnI and cTnC. To test this, computational alanine scanning was performed to calculate the change in the binding free energy (ΔΔG) caused by alanine substitution of all residues involved in the intermolecular interface contacts of the sTnI-cTnC structure (Fig. 4C). In cTnC, alanine substitution of E19 was found to contribute most to the ΔΔG (ΔΔG E19A = 0.91 kcal/mol). However, the greatest impact on the binding free energy among all residues engaged at the cTnC:sTnI interface was observed with the sTnI H132A substitution (ΔΔG H132 = 2.46 kcal/mol). This thermodynamic analysis supports the conclusion that alanine substitution of sTnI H132 was the most effective mechanism of reducing the binding free energy at the interface of the calcium-saturated sTnI-cTnC structure.

Interatomic distances were calculated between the cTnC E19 (OE2) and residue 132 of either WT sTnI (ND1) or sTnI H132A (CB) across a 40-ns atomic resolution MD simulation [Fig. 5, A and B; Suplemental Video S3 (sTnI H132A:cTnC)]. These data show that the WT sTnI-cTnC structure retained a highly stable interface mediated by the high-energy intermolecular salt bridge interaction between His132:E19. In contrast, there was a marked increase in the distance between cTnC E19 and sTnI residue 132 in the alanine-substituted structure, indicating a significant structural interface separation and destabilization of the complex (Fig. 5B). Positional root mean square fluctuations (RMSF) of all residues during the 40-ns MD simulation also indicate structural untethering within the sTnI H132A switch arm compared with WT sTnI (Fig. 5C). These MD simulations were replicated with a different randomly assigned starting velocity, and the results confirmed our initial observation (Supplemental Video S4). Functional data support these in silico studies. A recent report by Westfall and Metzger (112) has shown that sTnI H132A significantly reduces the calcium sensitivity of the myofilaments (pCa50) and enhances the relaxation performance of myocytes similar to the functional properties observed with myocytes expressing mammalian cTnI.

Given our initial assumption that sTnI recapitulates ectothermic cTnI, the molecular modeling data indicate that there was evolutionary pressure in the mammalian taxa to favor cTnI A164 in order to reduce the binding free energy (ΔΔG) at the interface of cTnI and cTnC. This was accomplished by eliminating the key high-energy electrostatic interaction between TnI and cTnC (H132-E19) making H4 and the COOH-terminal domain of TnI molecularly untethered. Together with published functional data (112), the phylogenetic and molecular modeling data here indicate that the modification in the switch arm of cTnI between amphibians (histidine) and mammals (alanine) significantly reduced myofilament calcium sensitivity through enhanced relaxation of cTnI in the mammalian cardiovascular system. The proline residue observed in platypus cTnI accentuates this point. Proline is the archetypal helix breaker and is also nonpolar. At the nucleotide level, the codon encoding proline may have acted as an evolutionary intermediate, as
in the following mutational codon transition: CAC (Histidine codon observed in anuran and saurian species) nonsynonymous mutation to CCC (Proline codon observed in platypus) nonsynonymous mutation to GCC (Alanine codon observed in mammals most closely related to platypus such as opossum and some other mammals) synonymous mutation to GCT (Alanine codon observed in all other mammals including humans). Thus we posit that, in terms of structural, functional, and molecular evolution, the platypus proline at codon 164 is an excellent intermediate to enhance myofilament relaxation of cTn through disruption of the histidine-glutamate salt bridge predicted to occur in fish and anuran species.

The hypothesis that this histidine vs. alanine residue in the switch arm of cTnI is important in the evolutionary divergence of ectotherm vs. mammalian physiology could be tested by an alanine-substituted cTnI into fish myocytes. If our hypothesis is correct, then an alanine residue in cTnI of fish (e.g., zebrafish) would markedly diminish calcium sensitivity of the myofilaments in vitro and would likely not be sufficient to support normal physiological requirements of that organism in vivo. Alternatively, if this model is correct, expression of ectothermic cTnI in mammalian myocytes would increase calcium sensitivity of the myofilaments.

Increased myofilament calcium sensitivity in nonmammalian chordates. Previous reports have shown that residues at the NH$_2$ terminus of cTnC are required for differentially modulating the calcium binding affinity of TnC between endotherms and some ectotherms (Supplemental Fig. S6) (32, 33, 35). It has been suggested that, in at least one ectotherm species, trout, four residues (Grp sequence: N2, I28, Q29, D30) are responsible for decreasing the energy barrier required for conformational changes in the NH$_2$ terminus of TnC, which increased calcium binding affinity at EF hand site II (32, 33, 35). In endotherms (birds and mammals), these residues have been converted with high conservation to D2, V28, L29, and G30 and markedly reduce the calcium binding affinity of cTnC (33). In mammals, failure to modify these residues results in cardiomypathy. Studies have shown that a single amino acid substitution, L29Q, in human cTnC results in hypertrophic cardiomyopathy the underlying cause of which may be a significant alteration in myofilament calcium sensitivity, although an alternative possible explanation involving potential alterations in the PKA-dependent phospho-cTnI:cTnC engagement has also been suggested (6, 44, 61, 85). Other pathogenic peptide deviations in mammals associated with increased myofilament sensitivity to activating calcium have been found in SNPs within other regions of cTnC as well as cTnT (9, 28, 55, 72, 73, 79, 81, 97, 98, 116). Together these studies indicate that evolutionary selective pressure to modify cTn occurred within the adult mammalian heart to reduce myofilament calcium sensitivity to an extent markedly below that in ectothermic chordates.

Enhanced relaxation of cardiac troponin required for mammalian heart function. There is a significant body of literature providing comparative analysis of cardiovascular function between chordate species (11, 17, 29, 42, 43, 53, 88, 89, 94, 104, 105). These studies reveal important details about the differences in physiological requirements that necessitate unique mechanisms of cardiovascular support among chordate species. The present study contributes to this knowledge base by providing the first integrated model for molecular evolution of the cardiac troponin complex. These data show that cTn modifications are critical for enhanced relaxation performance of the thin filament regulatory complex and may have evolved to accommodate unique demands for myofilament function of the adult mammalian heart (Fig. 6).

Compared with adult mammalian myofilaments, the expression of TnI isofoms with inherently high calcium sensitivity in the developing hearts of a range of chordate species (amphibian cTnI and bird/mammalian ssTnI) (25, 68, 84, 92, 113) indicates that heightened calcium sensitivity and pH resistivity of the myofilaments may be important for heart development in chordate species (Fig. 6). Several studies support this assertion. First, the mammalian fetal heart has high lactate levels and relies on glucose metabolism (64). Second, there is evidence that the critical cardiac developmental transcription factor Nkx2.5 is regulated by hypoxia-inducible factor (HIF)-1α (74). Third, immature sarcolemal reticulum and calcium cycling in the developing heart require heightened sensitization of the myofilaments to establish sufficient nascent heart pump function (31, 86).

However, in the adult heart, requirements for proper regulation of cardiac output are markedly different between chordate species. Numerous factors intrinsic and extrinsic to the heart affect the efficiency of heart rate and stroke volume to sustain cardiac output. Cardiomyopathic SNPs identified in all members of the human heterotrimeric cardiac troponin complex have been found to cause pathologies characterized by a marked increase in myofilament calcium sensitivity (19, 20, 26, 44, 46, 61, 87, 106). Together these studies show a strong relationship between cardiomyopathic SNPs in tropinin and heart failure phenotypes including diastolic dysfunction, arrhythmias, and sudden cardiac death.

In the context of chordate evolution, our current understanding of fish and amphibian physiology supports the need for heightened myofilament calcium sensitivity with pH resistivity compared with mammals (62, 88–90, 92). Similar to cTnI, other proteins have shown analogous adaptations in these species. For example, hemoglobin in many fish has an exacerbated Bohr effect known as the Root effect, an extreme form of pH-sensitive Hb O$_2$ binding (10, 14, 82). On the part of regulating cardiac contractility during hypoxia, modification in intracellular pH in response to temperature shifts, termed α-stat regulation, is thought to maintain cardiac myofilament function in ectothermic and poikilothermic species by keeping the fractional dissociation of histidine imidazole groups constant (16, 36). The NH$_2$-terminal NIQD motif of cTnC and the switch arm peptides of cTnI seen in fish and amphibians may also provide a substrate to support the required cardiovascular demands of ectotherms (Fig. 6).

The observation that pathogenic peptide deviations [either naturally occurring SNPs (20, 61) or in situ protein mutagenesis (32, 112)] cause cell physiological effects including hypersensitive myofilaments indicates that, compared with other nonmammalian chordate phylogenies, reduced myofilament sensitivity to activating calcium was a necessary adaptation within the mammalian adult heart. Phylogenetic alignment data show the emergence of protein kinase (PKA and PKC) phosphorylation sites at the NH$_2$ terminus of TnI (S23, 24) at the level of amphibians (Supplemental Fig. S2). One of the major purposes for protein kinase phosphorylation of these residues is to reduce the calcium sensitivity of the myofilaments in coor-
Coordination with changes in heart rate (56, 70, 96, 118). Although covalent modifications of TnI emerged prior to mammals, this mechanism for modulating myofilament function was evidently insufficient for mammalian cardiovascular function. As a consequence, this study supports the hypothesis that in the adult mammalian heart coordinated residue modifications affecting the key regulatory interaction between cTnC and cTnI enhanced the intrinsic relaxation potential of troponin independent of covalent modifications.

Structure and function data indicate that, in endotherms, coordinated residue modifications occurred in cTnC to markedly decrease the calcium binding affinity at EF hand site II (32, 33, 35). However, among all mammalian sequences identified, this study provides new evidence of a critical hydrophobic and nonpolar (primarily alanine) moiety substituted for histidine in the switch arm disrupting a strong intermolecular electrostatic interaction at the interface between cTnI and cTnC in the calcium-saturated state (Fig. 6). In silico MD alanine...
scanning simulations shown here together with recent functional data (112) support a model that, only within mammals, this His to Ala substitution reduced myofilament calcium sensitivity by making cTnI molecularly untethered from cTnC, thus reducing the thermodynamic potential at the intermolecular interface (ΔΔG). These data also indicate that loss of a therapeutically effective pH-responsive histidine side chain in cTnI (22, 75, 76) was necessary to reduce myofilament sensitivity to activating calcium (112). As such, this study makes use of disease-causing mutations in troponin as they provide insight into the evolutionary physiological adaptations of the mammalian cardiac myofilament. The acquisition of an alanine moiety in the switch arm of TnI with coordinate residue modifications in TnC reveals the basis for reduced cardiac myofilament calcium sensitivity in mammalian species.

ACKNOWLEDGMENTS

We thank the labs of the University of Minnesota Lillehei Heart Institute for questions oriented toward the phylogenetics of cardiac troponin that stimulated this study. We thank Nobuobe Ngandu, Dr. Cathal Seoighe, and Dr. Simon Frost for assistance with interpretation of selective pressures on the cTnI molecule. We thank Dr. Robert Kulathinal for his expertise in compensatory pathological mutations in evolution. We also thank the University of Minnesota Supercomputing Institute for providing the computational resources. Present address for N. J. Palpant: University of Washington, Center for Cardiovascular Disease and Regenerative Medicine, 815 Mercer, Seattle, WA 98117.

GRANTS

This work was supported by the National Institutes of Health (ROI-HL-059731 to J. M. Metzger and ROI-GM-076121 to A. Y. Ouniffiev), the American Heart Association (0715632Z to N. J. Palpant), and the University of Minnesota Center for Drug Design (Y. Y. Sham).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

REFERENCES

35. Gillis TE, Moyes CD, Tibbits GF. Sequence mutations in teleost cardiac troponin C that are permissive of high Ca2+ affinity of site II. Am J Physiol Cell Physiol 284: C1176–C1184, 2003.
MOLECULAR EVOLUTION OF CARDIAC TROPONIN


